

Opportunity for Europe's science

One of the best, but smallest, of Europe's research organizations wants to be bigger. Not before time. But will those who hold the purse-strings agree?

NEXT week's General Assembly of the European Science Foundation (ESF) at Strasbourg could be crucial for the future of basic science in Europe. ESF's members, more than 50 research councils from 19 countries (with Iceland joining next week), will there have to take a view of the scale on which ESF functions. Huge sums are not at stake. The annual budget of ESF now is less than FF14 million, or less than \$2 million. For the time being, the expansionists' plans would merely increase the scale of operations threefold. The choice is between the present shoestring and one only a little thicker. But there are two reasons why the issue is more important than that. First, ESF's budget is highly geared: each franc spent may be matched by 10 francs spent by member organizations. Second, the new activities on which ESF wishes to embark would give it, for the first time, an executive (if still small) role in the conduct of European collaboration in science. It thoroughly deserves one.

Here is why. ESF is catholic in the best sense. It follows the European mainland's broad understanding of what science is. The humanities and social sciences stand alongside nuclear physics and the exploration of interplanetary space. But ESF is deliberately not restricted geographically to the European Economic Community (EEC), but covers Scandinavia, Switzerland and Austria and even Turkey. People from Eastern Europe (and even China) have taken part as individuals in particular programmes. ESF now claims, rightly, that a nongovernmental organization such as itself is well placed to bring about more durable collaboration. Indeed, ESF seems to have been talking informally to two East European countries, but says that no "single dramatic initiative" is in prospect. ESF is certainly better placed to make a research bridge with the East than any other European institution. Will the tumultuous events of the past few weeks persuade next week's assembly to throw caution to the winds?

Success

As a club for European research councils, ESF has 16 successful years to boast about. It is organized to identify potentially fruitful fields for research collaboration, which are then carried through by groups of individual researchers supported, when necessary, by the research council members. The European Geotraverse Project, a coordinated programme (now almost finished) of deep seismic soundings from north to south in Europe (with a

projection into North Africa) is a good illustration. ESF launched the project in 1983 at the urging of interested geophysicists, since when research councils electing to participate have paid for the direct costs of coordination (more than FF1 million this year) and have spent much more on research grants to support the actual work. The general opinion is that the implications of the data gathered will be more valuable than would have been the sum of the several uncoordinated parts. Much the same can be expected of the programme, likely to be approved next week, on the genetics of psychiatric illness, recent disappointments notwithstanding (see page 222). The plan is to harness resources on a European scale in the search for families likely to be useful in the search for genes implicated in psychiatric illness.

ESRS

But ESF has other feathers in its cap, most conspicuously the European Synchrotron Radiation Source now being built at Grenoble. The case for that was first made by an ESF committee. The funds were prised out of reluctant governments (chiefly West Germany and France) by ESF diplomacy. It is absurd that so little credit should have stuck to ESF for its midwife's role. This project in particular would have been carried through in half the time (seven rather than 15 years) if ESF had had more executive responsibility. ESF now says it wishes to play a more direct part in projects such as these. It could do worse than cut these new sharper teeth in the European VLBI Consortium's plan for the more effective coordination of data gathered by Europe's radiotelescope, still orphaned for lack of funds (see *Nature* 337, 301; 1989).

What else is there to do? In the past few years, ESF has encouraged several successful informal networks of laboratories and individuals in fields as different as polar science and the use of longitudinal studies in childhood development. Again the principle is that ESF uses its own money to tell whether proposed networks will meet a need, often by organizing a sample workshop or two, whereupon administrative and operating costs are met by the participants. Now ESF wishes to go further — first by using its own money to organize research conferences and then by administering a scheme for up to 500 three-year postdoctoral fellowships a year. In a sharp break with previous practice, ESF would like to spend its own funds on these ventures, which means a bigger budget. Expli-



citly, ESF is saying that it has won the right to be regarded as the natural executive agent of its members for transnational collaboration on a European scale.

The argument is sound, but will it be heeded? Two issues are bound to dominate next week's debate: how promising are the new schemes, and can they be afforded? The underlying objectives are similar — to find ways of giving Europe's scattered academic communities a sense of being part of a critical mass. ESF hopes to be given funds next week to run a two-year pilot programme of conferences, but the case for the potentially more expensive postdoctoral programme is even stronger, if only because there is at present no general means by which people embarking on careers as independent researchers can be helped to move from one European academic centre to another. The underlying objective is to create within some European geographical context a sense that the academic community is a well-integrated whole, untrammelled by national or state boundaries. Until that happens, European science will not make the best of its potential and will remain at a disadvantage with North America.

But there are also difficulties. In breaking this new ground, ESF will, for example, for the first time encounter the too-familiar corrosive internal arguments about 'fair shares' that plague most international organizations: why should members from one country contribute when the immediate beneficiaries may be from elsewhere? The complaint mistakes the objectives, which are communal, not sectional. Specifically, postdoctoral fellows from one country who serve their time in a second and then become full-time academics in a third are only a 'loss' to the first two if the European academic community remains as balkanized as it is at present, which amounts to saying that neither ESF's plans nor other initiatives in the air will have much effect. That is a counsel of despair.

There are also practical difficulties; most postdoctoral fellowships are linked with research grants, for example, especially in the natural sciences. A postdoctoral fellow cannot work effectively without research funds, but universities are at present able to provide for only a few out of their own resources. But it would defeat the purpose of the new schemes if they were confined to fields in which research is relatively cheap, as in the social sciences. Uncertainties such as these, rather than the sheer cost of the postdoctoral scheme (which nevertheless could easily exceed the equivalent of \$30 million a year) will determine what happens next week. It is to be hoped that ESF's assembly will be ready at least to experiment.

What matters most of all is that ESF's members should have a clear idea of why they are in business at Strasbourg. ESF is by no means the only European institution for the encouragement of transnational science, but it is a distinctive one. It differs from the organs of the European Economic Community (EEC), for example, geographically and by its concern for basic rather than applied science. (The EEC should pay more attention to basic science, but that is another matter.) ESF is also concerned with quality

in research, not merely with getting things done. However much it grows, it will not in the foreseeable future be a big spender, but its record shows that its influence can far exceed the scale on which money runs through its hands. The opportunity, now, is to recreate the circumstances of the eighteenth century, when European scholarship was free and also European in the broadest sense — manageably so because it was so much smaller. The prize, the reinvigoration of European scholarship — could also be disproportionately handsome. □

Strength via adversity?

Having outlasted an administrative attack, the US fusion programme must not subside into somnolence.

THE only thing worse than a federal science administrator who knows nothing about science is one who does. That, apparently, is the lesson that physicists in the US magnetic fusion programme will be obliged to draw from the recent departure of Robert Hunter from the Department of Energy, where he had been director of energy research. In Congress and elsewhere, those who have been trying for decades to persuade tokamaks to burn deuterium and tritium had been complaining about Hunter's activities: he diverted funds from magnetic to laser fusion research, tried to stall construction of the Compact Ignition Tokamak and, worst of all, refused to take physicists at their word when they assured him the magnetic fusion programme was proceeding as planned towards commercial energy generation.

Hunter was, no doubt, politically a clumsy man. He shifted money from one budget category to another without consulting the appropriate Congressional committees. He further irked Robert Roe, the chairman of the House Committee on Science, Science and Technology, by observing in a departmental memorandum that Princeton was not the automatic choice for the site of a new tokamak because of political concerns over radioactivity; Roe, coming across this note at a fractious hearing, was in no mood to be told by a federal administrator what was or was not a political issue in his home state.

There will be many sighs of relief and little lamentation at Hunter's going. But before the tokamak community congratulates itself too much, they should ponder something: for the first time since the oil crisis of the early seventies, fusion energy was the subject of a noisy political debate. A dozen congressmen grappled intently with plasma quality factors and confinement times. In fighting off Hunter's attack, scientists who had languished under level funding for fifteen years found themselves being praised by some influential voices. With a more timid and less engaging character in Hunter's place, the magnetic fusion programme may well return to the political backwaters. Physicists had better make sure that this does not also mean a return to the financial neglect and sloth of the last decade. □

Nuclear installations not the cause of cancer?

- Risks similar at rejected and chosen sites
- Unidentified risk factor may be involved

Washington

RESEARCH published in Britain last week shows that the mortality rate from leukaemia and Hodgkin's disease in young people is higher not only near existing UK nuclear power stations but also near places being considered as future power station sites. This further tips the balance of the evidence away from the theory that radioactive discharges are behind the increases.

Epidemiological research in Britain, the world leader in studies of the health effects of nuclear installations, has so far been inconclusive in attempts to prove or disprove a link between radioactive discharges and an increase in the incidence of leukaemia and Hodgkin's disease in young people living nearby.

After examining the evidence last year, the Committee on Medical Aspects of Radiation in the Environment, an independent commission, concluded that some unidentified feature of Britain's nuclear reprocessing plants at Sellafield in Cumbria and Dounreay in Scotland was the most likely cause of the increased incidence of leukaemia, but could not attribute the increase to radioactivity in the area because the measured levels were so low.

Leo Kinlen of the Cancer Research Campaign's epidemiology unit at the University of Edinburgh later found evidence to support the theory that the increased leukaemia incidences might result from viral infections. His theory proposes that a rapid influx of population in the vicinity of a nuclear installation, usually located in an isolated area, brings an infection to which the indigenous population has not evolved resistance.

In an attempt to decide whether the increased cancers are caused by radiation from nuclear installations or by some other contingent feature, Paula Cook-Mozaffari, Sarah Darby and Richard Doll of the Imperial Cancer Research Fund Laboratories in Oxford compared mortality data from areas near two Central Electricity Generation Board (CEGB) nuclear power stations and six other sites being considered by the CEGB for the construction of power stations. They reported in *The Lancet* last week that the increased risk of mortality due to leukaemia and Hodgkin's disease in young people was "strikingly similar" in both groups.

Although they declined to suggest any reason for the increased mortality, they

suggested that these areas might share "important unrecognised risk factors" other than environmental radiation pollution. But they also point out that there had been no sudden population influx in the potential sites, which argued against the viral infection theory.

Evidence is now mounting, says Darby, against the theory that radioactive discharges are responsible for the increased cancer risk, although at Sellafield, home to one of Britain's two nuclear reprocessing plants, the discharges and the increased cancer risk are much higher and

the issue is still open. At Dounreay in Scotland, the site of Britain's other reprocessing plant, the reported increased leukaemia incidence is also significantly higher than in the areas in this study.

Kinlen says the new results are consistent with the theory that in isolated coastal areas, which would be suitable sites for nuclear installations, the pattern of human interaction is different from that in the average large inland town and could explain the increased risk for certain types of cancer. He adds that "it is possible, if not probable" that the new results are recording the same effect he saw on a smaller scale and predicts that a similar result would be found in any rural coastal area, regardless of whether it was a potential site for a nuclear power station.

Christine McGourty

Cancer near potential sites of nuclear installations, by P. Cook-Mozaffari, S. Darby & R. Doll, *Lancet* II, 1145-1147 (1989).

NUCLEAR POWER

Privatization spells end of a nuclear dream

Washington

THE future of the nuclear industry in Britain is in doubt after last week's announcement by the government that it was abandoning plans to build three new PWRs (pressurized water reactors). The reversal of the plans to expand the nuclear industry is an about-face by the Conservative government under Prime Minister Margaret Thatcher, a staunch advocate of nuclear

power, but was inevitable given the government's determination to privatize the electricity industry and the refusal of the soon-to-be-privatized local electricity boards to carry through the £7,000 million PWR programme and face the huge costs of nuclear waste disposal and of de-commissioning old power stations.

Strong opposition to nuclear power forced a long and controversial public inquiry into the safety and design of Britain's first PWR at Sizewell in Suffolk. The plans were approved 18 months ago and construction is under way. The subsequent inquiry into a second PWR, at Hinkley Point C in Somerset, has been running for more than six months and was drawing to a close, but will now be abandoned along with plans for another two PWRs at Sizewell C, and Wylfa B in Wales.

Opponents of nuclear power scored a major victory during this inquiry when the

Central Electricity Generating Board, which has always promoted the low costs of nuclear power, was forced to disclose figures showing that its coal-fired power stations are cheaper.

In an attempt to ensure that nuclear power was not neglected after privatization, the government had stipulated that by the year 2000, 20 per cent of the electricity demand must be met by non-fossil



Former director of the UK Atomic Energy Authority, Lord Marshall may not be the man to run a coal-fired CEGB.

fuel sources to maintain fuel diversity. A new private company, National Power, was to be responsible for nuclear power generation. But its chairman-designate, Lord Marshall, is confidently expected to resign; and now the 10 existing Magnox reactors and 7 AGRs (advanced gas-cooled reactors) will be retained in a new publicly owned company.

Christine McGourty

Testing time for Caltrans

San Francisco

A SIMULATED earthquake on a portion of intact freeway and an elaborate computer model of the San Francisco Bay Bridge are among the studies likely to be adopted by California officials in an effort to understand what happened during the 17 October Loma Prieta earthquake. At a meeting scheduled for late last week between representatives from Caltrans, the state transportation department, and University of California engineers, several proposals for roadway studies were due to be debated, but Caltrans spokesman Jim Drag confirmed that the simulated quake and the Bay Bridge studies seemed likely to be approved.

Also last week, California governor George Deukmejian signed 24 earthquake relief measures, including a sales

tural engineer at the University of California at Berkeley, is to test proposed strengthening techniques for other San Francisco Bay Area double-decker freeways. Three similar roadways in San Francisco that were damaged in the earthquake are still closed.

In the second test, which will alternate with the hydraulic-jack experiment, a pair of vibration generators will be attached to the same freeway section. Attached to the upper deck, the twin 'shakers' rotate heavy weights in opposite directions at a synchronized frequency, so that the centrifugal force is focused in a specific direction along the roadway. The force, of the order of 10,000 pounds, will not be great, but by repeated application at a steady rate it will test the structure's endurance against vibration.

Accelerators attached to the freeway will allow engineers, knowing the mass of the deck, to measure the forces developed in the supporting columns. "If we know the total force in the structure and how it displaces under that force, we can develop a good computer model of the resisting mechanism", says Moehle.

The tests will be finished by Christmas, when the entire freeway span will be

Looking south on Interstate 880 on the Tuesday evening. (AP). demolished. Because the tested section will not be "failed" in its original condition, engineers will have no direct measure of how the Cypress section really collapsed, but the investigators hope that demolition crews will haul a few columns away for further study in the laboratory.

The highway test involves an intact section of Interstate 880 in Oakland, just south of the Cypress Viaduct section which collapsed during the upheaval. Apart from a firmer foundation, the section is almost identical to the portion that collapsed. The prevailing theory is that the upper deck failed because the supporting columns and the joints that bind them with the lower deck had not been adequately reinforced. Two tests will examine these and other ideas.

In one, hydraulic jacks mounted on steel frames under both sides of the roadway will be used to simulate the sideways movement of the 7.1-magnitude quake. An array of electronic instruments and cameras will measure the displacement and strain developed at specific points, and record the physical damage.

The overpass will first be tested in its present condition, after which engineers will reinforce the columns with steel beams, and then slowly bring the freeway to the failing point, which might take as much as three million pounds of force. The objective, says Jack Moehle, a struc-

tural engineer at the University of California at Berkeley, is to test proposed strengthening techniques for other San Francisco Bay Area double-decker freeways. Three similar roadways in San Francisco that were damaged in the earthquake are still closed.

The Bay Bridge study will be confined mainly to laboratory simulations and computer modelling. Faculty and students from the structural engineering programme of UC Berkeley's civil engineering department have already been amassing damage information, from notebooks, photographs, videotapes and sketches. They plan to supplement this data with other input to develop a detailed computer model of the bridge that extends even to its underlying soil.

"Our main goal is to make sure our computer programs — simulations, modeling, all the tools that we have to do the analysis — are capable of predicting events like this and damage like this", said Abolhassan Astaneh, a UC Berkeley assistant professor of civil engineering. These studies, he hoped, will enable engineers to recommend specific retrofitting plans and other modifications, so that the bridge can withstand "the big one".

Robert Buderl

Federal policy makes trouble

Washington

MANY examples of "waste, fraud and abuse" in the federal government can be attributed to problems in developing software, according to a report released last week by the US House of Representatives Committee on Science, Space and Technology. On everything from budgeting to intellectual property rights, it says, government policies "have congealed over time in a manner almost perfectly designed to thwart the development of high-quality software".

Although no single action will instantly solve the government's software woes, the committee staff recommends that a working group on improving software development be set up in the president's Office of Science and Technology Policy with the aim of making long-term changes in the government procurement system.

At present, government contracts specify in "excruciating" detail exactly how a system will look at delivery several years in the future, so when the inevitable changes are introduced, software quality is adversely affected and development costs increase.

Reducing overall costs would require spending more time and money in the design stage, when it is much cheaper to correct an error, than in the maintenance phase. But this strategy is discouraged by the process for obtaining congressional approval, which requires initial costs to be minimized.

The report, stemming from an investigation into faulty software in a radiation therapy machine which caused several deaths, also calls on the National Institute of Standards and Technology to play a greater role in improving software evaluation. At present there is no infallible method for assessing quantitatively the safety of software used in, for example, aircraft flight controls or nuclear reactors, in which software flaws can put lives at risk. The federal government's "software woes" are exacerbated by the staff crisis created by its inability to compete with the private sector in attracting software specialists.

Computer security and ethics also came under scrutiny last week in the House subcommittee of criminal justice, which is considering two bills to increase the penalties for damaging hardware and software as a deterrent to hackers. But researchers and legal experts are divided over the need for new legislation, definitions of "authorized access" and how to increase computer security without restricting free exchange of information. Action on the bills is being delayed pending the outcome of the trial of Robert Morris, the Cornell University graduate student whose 'virus' disabled thousands of computers last year.

Christine McGourty

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

COLD FUSION

Official thumbs down

Washington

To no one's surprise, the panel set up by the Department of Energy (DoE) to assess cold fusion has sent in a negative report. An interim opinion released earlier this year by the panel, a group of 23 electrochemists, nuclear physicists and materials scientists assembled by the DoE's Energy Research Advisory Board (ERAB), acknowledged that there were some odd and unexplained results in the air but recommended against any special support of cold fusion experiments (see *Nature* **340**, 174; 1989). The panel's final report expresses the same verdict a little more forcefully, but the ERAB meeting at which the report was approved and passed on to DoE Secretary James Watkins did reveal some differences of opinion on the significance and credibility of the handful of 'positive' results still extant.

Summarizing the final report, John Huizenga, a professor of physics and chemistry at the University of Rochester and one of the ERAB panel's co-chairmen, showed little patience with those who have continued to claim that there is at least a little evidence for cold fusion of the heat-producing variety. None of the calorimetry, he declared, was of good enough quality to be believed, and in any case calorimetry alone would never prove the case. Huizenga was particularly unhappy with those who argued that because the claimed excess heat could not possibly be of conventional origin it had to be nuclear, despite measurements which showed unequivocally that no fusion products were being generated.

Discounting the calorimetry, the only anomalous claim still needing an explanation was that of Kevin Wolf of Texas A&M University, who repeatedly finds tritium in his electrochemical cells. But new evidence from Wolf, discussed at last week's ERAB meeting, argues against a nuclear origin. Tritium produced by deuterium fusion comes with an energy of 1 MeV and is accompanied by a 3 MeV proton. Energetic tritons will react with deuterium, either in the palladium electrode or in the electrolyte, producing secondary neutrons, and the protons can excite some palladium nuclei, which give up a 500 keV photon on de-excitation. In his latest experiments, Wolf has searched for both these secondary signatures and found nothing. His tritium, therefore, must have some kind of low-energy origin.

Nevertheless, Wolf has also searched minutely for sources of contamination, and found nothing there either.

This puzzling result, together with a few calorimetric measurements that he found somewhat convincing, led recent Nobel prizewinner Norman Ramsey of Harvard University (see *Nature* **341**, 556; 1989),

the panel's other co-chairman, to conclude that the case against cold fusion was not closed. At his insistence, a preamble was added to the report's conclusions which notes that "it is not possible to state categorically that all the claims for cold fusion have been convincingly either proved or disproved".

Huizenga felt that this preamble "weakened" the report, but Ramsey and some others felt that the report would undermine itself if it looked too dogmatic in its dismissal of the evidence.

On the subject of the low-level fusion process espoused by Steven Jones of Brigham Young University, the panel agreed that more work was needed, although here again Huizenga, noting that

EAST GERMAN EXODUS

Problems remain for those who stay

Munich

EAST German researchers now living in West Germany are happy but sceptical about the events in East Germany that culminated in last week's opening of the Berlin Wall. No one disputes that the opening of the borders is a momentous and irreversible step, but many doubt that the effects of 40 years of Communist repression can be wiped away overnight: the repressive system is more deeply entrenched in East Germany than, for example, in Hungary, which has also enacted radical reforms this year.

Under the old system, said a number of researchers, Party members and other desirables were given the opportunity to travel abroad, publish in foreign journals and make progress in their scientific careers, but those who lacked an ideological pedigree were cut off from the West and forced to engage in political activities against their

West Germany continued to welcome all those who came over from East Germany on Friday, even though the capacity of the refugee camps had nearly been exhausted. Most of the refugees who had arrived before the Wall came down seemed to be in practical, not academic, professions, although an accurate tally was made impossible by the large numbers. But organizations involved in helping East German researchers and academics get on their feet in West Germany said they would need much more money to continue the process.

Uta Paffhausen of the Education Ministry reported that the number of applications for "adjustment aid", given to help East German emigrants find academic posts, had risen to about 50 in the past two months. The grants are provided by the ministry as well as by the Otto Benecke Foundation. The ministry also administers a programme to provide fellowships for refugees whose training is unsuitable for working in West Germany. The two programmes were to provide a total of DM 24 million (\$13 million) in 1990 when the refugee total was still expected to be about 60,000. But by 10 November, the number had topped 500,000, including 200,000 East Germans and another 300,000 ethnic Germans from Poland and the Soviet Union. Photo: The wall near Checkpoint Charlie last Friday, seen from the East (AP).

all the claims were of marginal statistical significance, seemed less impressed than some of the other panel members that there really was a phenomenon to be explained.

Although he showed some sympathy for those still making claims for cold fusion, Ramsey also said that if the results on which the claims were based came from his laboratory, he would keep quiet and do more experiments until he understood better what was going on. The ERAB report accepts that people are free to submit experimental proposals to granting agencies in the hope of obtaining support for cold fusion experiments, but recommends that only if experiments are likely to yield useful results — closed-cell calorimetry with simultaneous nuclear monitoring is deemed essential — should support be given.

David Lindley

will. The pessimism brought on by such obstructions caused these people, like so many others, to escape.

Those who tried to leave by legal means found that as soon as they submitted visa applications, they were forbidden to work in their laboratories. One said he was reinstated only after an invitation to a scientific meeting from a prominent West German research institution, in which he now has a job. But the only way to find out about such opportunities is to have personal and professional contacts, said another.

Researchers said after the recent lifting of travel restrictions that they were waiting to see if the changes are large enough to encourage development of a system based on personal achievement and not privilege. Even so, East German research laboratories will still lack the money to order basic supplies, let alone to subscribe to foreign journals.

Steven Dickman

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Politicians still only talking

Boston

IN a week marked by vocal debate over global climate change, a group of representatives from private foundations gathered quietly with scientists and international security experts in Cambridge, Massachusetts, last week to explore the international political fallout from anticipated changes in the environment.

While British Prime Minister Margaret Thatcher called upon the UN General Assembly to lead the way in worldwide action, and environmental policy-makers from 68 nations gathered in the Netherlands failed to agree on international limits on carbon-dioxide emissions, those meeting in Cambridge tackled social and legal issues such as the prospect that drastic climate change could produce 'environmental refugees' and the idea that environmental concerns could become so crucial that countries would be justified in overriding the sovereignty of others to enforce protective measures.

At the Cambridge meeting, which was sponsored by the American Academy of Arts and Sciences, the presence of representatives of some of the largest private US foundations signalled a mounting interest in global change research. Although no new financial commitments were announced at the meeting, several foundations, including the John D. and Catherine T. MacArthur Foundation and the Social Science Research Council, have already earmarked funds specifically for research on global environmental change.

But dissent over political choices overshadowed scientific uncertainty. The scientists present agreed that precise and consistent models of global climate change are decades away, but nonetheless presented a fairly uniform and troubling consensus about the threats posed by ozone depletion and trapped greenhouse gases. It was the social scientists and security experts who clashed most severely in their assessment of the political implications of the issue.

Some, like George Rathjens, professor at the Center for International Studies at the Massachusetts Institute of Technology, argued that the economic costs involved in curbing emissions may be too high to allow immediate action, and that the costs "diminish the prospects for international agreements". But others declared that "we know enough now" to justify a push for international cooperative measures — even coercive measures.

Speaking unofficially on behalf of the Brazilian government, Eunice Durham, an anthropologist at the University of São Paulo, bristled at even the notion of international coercion or pressure to reduce global warming. In the view of the Brazilian government, said Durham, "military

action [would be] justified" if the nation's sovereignty over the Amazon rain forest is in any way challenged. Furthermore, she added, deforestation "is not going to be stopped unless there is a viable alternative development plan for the region". Similar divisions and disagreements emerged at the international conference in the Netherlands. Despite President George Bush's claims of the need for a "war against global warming", the United States, together with Japan and the Soviet Union, led the way in thwarting the endorsement of a specific timetable for

curbs on emissions of carbon dioxide. Most of the 68 countries present at the Netherlands environmental conference wanted to freeze allowable emissions of carbon dioxide by the year 2000.

William K. Reilly, head of the US Environmental Protection Agency, stressed the fact that the United States agreed in principle to the need for international emission levels. But many critics in the United States and elsewhere were disappointed at the outcome of the Netherlands meeting, and criticized the United States for its deliberate inaction. As one critic put it, the US position made Bush's strong statements about global warming look like "a lot of hot air". **Seth Shulman**

GLOBAL KNOWLEDGE

Americans and Russians mostly at sea

Washington

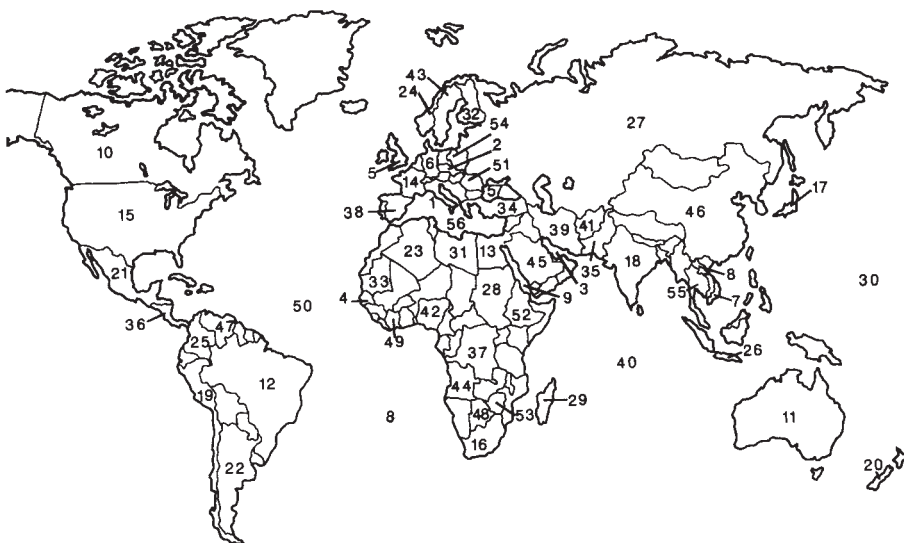
WHEN George Bush and Mikhail Gorbachev meet upon the Mediterranean Sea next month, they will be on equal terms in at least one respect: their respective countrymen will have an equally vague idea of where in the world they are. Asked to identify 16 locations on the map below, US citizens got an average of 8.6 correct, and Soviets did a little worse, scoring 7.4. Swedes and West Germans did best (11.6 and 11.2 correct identifications); the Japanese averaged 9.7 and the British 8.5.

The results of the survey, conducted by the Gallup organization for the National Geographic Society specifically to compare knowledge of geography in ten countries, have been taken in the United States as more evidence that the educational system, no matter how many brilliant scholars it may produce, is failing the general popula-

tion. For Americans, the most alarming result of the country-identification quiz was that the youngest age-group (18-24) did significantly worse than the rest of the population: for the other countries, younger people tended to do as well as or better than their elders. Last year, an international survey put US high school students near the bottom of the class in science (see *Nature* 332, 195; 1988).

But the Soviets, if the responses of 1,500 inhabitants of Moscow and Kursk reflect the knowledge of the population as a whole, have even more reason for alarm. Although 88 per cent of the Soviet sample had taken a geography course in school, and 61 per cent thought geographical knowledge a necessity, their achievements belied these ideals. Only 38 per cent of the Soviets knew where Afghanistan was.

David Lindley



LOCATE THE FOLLOWING: Write in the corresponding number that locates each place.
Answers on page 218. (Based on NGS questionnaire.)

- | | | | |
|-------------------|----------------|------------------|-----------------|
| — United States | — Canada | — Italy | — West Germany |
| — Soviet Union | — France | — Sweden | — Pacific Ocean |
| — Central America | — Persian Gulf | — United Kingdom | — Egypt |
| — Japan | — Mexico | — South Africa | — Vietnam |

AGENT ORANGE

'Not guilty' verdict challenged

Sydney

FOUR years after a Royal Commission adjudged that the widely used defoliant Agent Orange had caused no measurable or lasting ill-effects in soldiers in Vietnam, a group of Australian scientists has called the commission's findings unscientific. Eighteen biologists, chemists and immunologists from some of Australia's leading universities held a conference in Canberra last April under the title *Evatt Revisited*, in reference to the 1983 Royal Commission on the Use and Effects of Chemical Agents on Australian Personnel in Vietnam, headed by Justice Phillip Evatt. The proceedings of the conference were published last week, and include the charge that the commissioners simply duplicated large sections of the submission by Monsanto, the US chemical company that makes Agent Orange.

The Evatt Royal Commission was set up to determine whether Australian servicemen who served in Vietnam were more likely than others to suffer from cancer and whether their children had a higher probability of birth deformities. Since the commission released its findings in July 1985, it has been the target of complaints from the Vietnam Veterans Association of Australia (VVAA), from witnesses who testified before it and now from some of Australia's most senior scientists. Critics say, among other things, that the report is shallow and extravagantly worded, that scientific evidence was poorly interpreted and that VVAA witnesses were treated as "hostile".

TOXIC WASTE DISPOSAL

In 1987, the federal government ordered an inquiry into the findings of the Royal Commission, and concluded that the commission should not have come out with its "not guilty" assertion. The report maintained that the commission's findings and conclusions were sustainable but that it lacked credibility because of the manner in which the inquiry was conducted. But the new criticisms raised in *Evatt Revisited* are the first to complain about the commission's handling of the scientific evidence.

Ted Steele, a biologist from the University of Wollongong and one of the editors of *Evatt Revisited*, says that the commission's conclusions that exposure of civilians to Agent Orange had little or no effect on birth abnormalities was not warranted. He says that new evidence of birth abnormalities has come forward since the Royal Commission did its work, and believes that such findings "suggest that more cautious recommendations should have been forthcoming both from the Royal Commission and their scientific advisors". But Steele also has a more serious charge: he says that "large tracts of the report were plagiarized verbatim from the submission of Monsanto". But John Coombs QC, counsel assisting the commission, says Steele's allegation is "nonsense". He points out that it is common practice for a commission to incorporate appropriate parts of the submissions received into its report, and adds that "part of the VVAA submission was also reproduced".

Tania Ewing

US SECURITY

FBI interference with libraries criticized

Washington

FRESH concerns arose last week over interference by the Federal Bureau of Investigation (FBI) in the operation of academic and other public libraries. A year ago, the FBI came under attack for its long-standing Library Awareness Program (LAP), in which it tried to enlist the help of librarians in identifying foreign spies who might be seeking access to scientific and technical libraries (see *Nature* 336, 507; 1988). Now documents obtained by the National Security Archive (NSA), a non-profit Washington library and research centre, reveal that librarians who refused to help the FBI or publicly criticized LAP were subsequently investigated to see if the campaign against LAP was itself part of a Soviet plot.

The NSA, whose collection includes a variety of unclassified and declassified government materials, requested documents concerning LAP under the Freedom of Information Act, and then filed a follow-up lawsuit when the FBI failed to deliver. Under an agreement reached earlier this year, the FBI has now released 1,200 pages of internal documentation, blacked out in some places.

The FBI memoranda reveal that activities under LAP were continued through the past two years despite assurances from senior officials to the American Library Association (ALA) that the programme was put in abeyance after December 1987. FBI agents were, however, told not to refer to LAP by name because of the adverse publicity. And one internal FBI communication, dated 6 February 1989, notes that although 266 individuals had been subjected to "indices checks" since November 1987, only eight "references" were found. The purpose of these background checks, according to the FBI memo, was to "determine if a Soviet active measures campaign had been initiated to discredit the Library Awareness Program". The nature of the eight "references" turned up is not revealed.

But according to an FBI spokesman the ALA and the NSA are reading sinister motives into routine matters. Whenever anyone writes to the FBI on any subject, files are checked to see if the person has had any previous contact with the FBI, has been investigated by the bureau, has a criminal record, and so on. The "indices checks" referred to in the memo are simply searches through existing FBI files to see if a positive entry exists in any one of about 250 categories. The spokesman emphasized that these checks do not involve sending out agents to conduct personal investigations or questioning. Nevertheless, lawyers for the ALA are sifting through the 1,200 pages of FBI documents to see if there are any grounds for further action.

David Lindley

Commercial plans delayed

Cape Town

PLANS by a Cape Town company, Peacock Bay Environment Services, to build a toxic-waste disposal plant that would process imported as well as domestic waste have been put on hold by Gert Kotze, South Africa's minister of the environment. Kotze announced last month that the government would not consider licensing the importation of toxic waste before it had been presented with the results of a newly commissioned survey on domestic waste disposal. The survey will be conducted by the CSIR's Foundation for Research Development.

According to Sidney Sanders, a director of the company, the importation of foreign waste is essential for the economic success of the proposed R400 million (\$160 million) development.

He also claims that the plant offers "the first viable alternative to the inadequate way in which toxins are being disposed of in this country". But unease over the plans derives more from worries over the transportation of waste than its treatment at the

plant. A local toxic-waste transport company has indicated its reluctance to transport more waste by road, because the risk of accidents is too high, but Peacock Bay proposes to use the rail system instead.

The development, whose intended site is Alexander Bay, on the west coast 600 km north of Cape Town, is believed to be the first proposed under the new Environmental Conservation Act, promulgated by parliament in June of this year. The Act makes environmental impact assessment a legal prerequisite, and the Peacock Bay company says it has followed the procedures laid down by the act "to the letter", and has spent R750,000 (\$300,000) on an impact assessment study.

But local environment groups, spearheaded by the newly formed pressure group Earthlife Africa, have come out against the idea of importing toxic waste. Kotze, meanwhile, has accused Earthlife Africa of sowing suspicion about the government's integrity, but his department nonetheless appears to be sensitive to public reaction.

Michael Cherry

New law is overdue

Munich

A WEST German court last week ordered Hoechst AG to cease construction of a nearly completed plant that would use genetically engineered bacteria to make human insulin. The court ruled that it could not approve the licence granted for the construction of the plant because there is no legal basis for regulating the use of genetic engineering. The decision increases the already intense pressure on the government to pass a law regulating genetic engineering in factories and laboratories.

The court rejected the relevance of the West German Federal Emission Protection Act, which had been used as a legal basis for constructing at least one other facility using genetic engineering. The decision reversed earlier rulings by lower courts which had upheld Hoechst's right to build the plant despite legal challenges from local residents. Hoechst may appeal to the federal constitutional court at Karlsruhe for a reversal of the decision on constitutional grounds, but has not yet decided if it will do so.

The Hoechst plant would use the bacteria *Escherichia coli* to produce human insulin, as US companies have been doing for a decade. Hoechst has invested an estimated \$60 million in the plant and a spokeswoman said that the company intends to "fight further" for the use of genetic engineering in drug manufacture in West Germany. She called it "astounding" that there could be so much fuss over a technique that uses *E. coli*, which is found in the human intestine.

But the debate over genetic engineering here has been marked by the almost religious belief held by opponents that all gen-

etic engineering is hazardous. Although declining to rule on ethical or safety aspects, the court seemed to agree, saying that genetic engineering represented a "new dimension and quality" of technology with "risks for human beings and the environment that could not be properly assessed at this time . . . Only the democratically elected legislature" can decide how genetic engineering could be regulated.

The court's decision adds urgency to efforts to speed passage of a 'basic law' for genetic engineering that has been under discussion all year (see *Nature* 340, 85; 1989). The cabinet published its views on the latest version of the law on 9 November, the day after the court decision, preparing the way for discussion in the Bundestag this week.

It made some concessions to the *Länder* (states), for example agreeing to more public participation in the licensing procedure, but rejected the request of the *Länder* that they be allowed to carry out the provisions of the new law. The cabinet warned that this could potentially lead to a variety of interpretations of the law, which would be bad both for industry and for basic research.

But before the law can be put into practice, detailed regulations must be drawn up by the Health Ministry and the regulations attached to the law are "as important as the law itself" for basic research, said Ernst-Ludwig Winnacker, vice-president of the Deutsche Forschungsgemeinschaft (DFG). The effect on basic research "will depend on the wording", he said, and DFG, which is the main source of outside funding for basic research at universities, has volunteered to help draw them up.

Steven Dickman

ANTI-CLOTTING DRUGS

Approval for TPA competitor

Washington

THE approval of SmithKline Beecham's blood-clot-dissolving drug Eminase by the blood and blood products committee of the Food and Drug Administration (FDA) on 31 October puts the drug just one step away from being released for sale on the US market. Eminase, a complex of the already available thrombolytic drug streptokinase with human plasminogen, had been expected to win approval several months ago, but was held up for additional scrutiny by an FDA advisory committee amid reports of pressure from Genentech, which manufactures the rival drug Activase (tissue plasminogen activator, TPA).

David Stump, Genentech's director of clinical research, reported to the advisory committee that he had evidence of antibody production in response to the

plasminogen portion of Eminase.

But the results presented by SmithKline Beecham indicated no evidence of increased autoimmunity. Alan Wachter, manager of scientific information at SmithKline Beecham, emphasized that the approval of Eminase was unanimous. Although no price has yet been set, the company says the drug will be cheaper than existing thrombolytic drugs. Eminase has a longer half-life in the bloodstream than streptokinase, and can therefore be administered in 2–5 minutes instead of being infused over an hour. Activase has to be given over a period of three hours.

A large mortality study comparing the efficacy of Eminase, streptokinase and Duteplase (Borroughs Wellcome's TPA) is in progress and results are expected within the next few years. **Diane Gershon**

Political pressure means poor research

Washington

IN a report issued last week, the Institute of Medicine (IOM) called on the NIH to coordinate vigorous research programmes in basic reproductive biology and in the technologies of medically assisted conception in order to improve the practice of *in vitro* fertilization (IVF) and embryo transfer in the United States. Neglect of research in this politically sensitive area, the report says, is slowing the development of better methods of contraception, better techniques to preserve endangered species and more cost-effective methods of food production.

At present, embryo and fetal research require the approval of the government's Ethics Advisory Board (EAB). But the board has not been reconstituted since its charter expired in 1980. The IOM report complains that the eight-year-old ban on the use of federal funds for this kind of research means that more is now known about reproduction in commercially valuable animals such as cattle than in humans. Moreover, the clinical practice of IVF and embryo transfer is outstripping its scientific foundations.

After the renewal two weeks ago of the ban on research involving the transplantation of fetal tissue into humans (see *Nature* 342, 105; 1989), there is little optimism that any steps will be taken to revive the EAB. The IOM recommends that if politics paralyses the existing review mechanisms, then a nongovernmental organization, along the lines of Britain's Voluntary Licensing Authority, should be established to develop guidelines for embryo and fetal research.

But in addition to the political pressures, the report says that the lack of vocal advocacy groups to fight for major increases in federal support, the lack of communication between researchers in the different fields of reproductive biology and the lack of research material for experiments in humans and other primates all contribute to the slow progress in this field and to the low success rates of IVF and embryo transfer.

To improve the safety and effectiveness of the techniques used, the IOM suggests that the professional societies should establish a mechanism to collect and monitor data from clinical centres.

Christine McGourty

World map answers

Answers to NGS map questions (page 216). United States (15), Soviet Union (27), Central America (36), Japan (17), Canada (10), France (14), Persian Gulf (3), Mexico (21), Italy (1), Sweden (43), United Kingdom (5), South Africa (16), West Germany (6), Pacific Ocean (30), Egypt (13), Vietnam (7).

Small changes, big fuss

San Francisco

AFTER months of public debate, the California Board of Education last week approved new curriculum guidelines that give evolution a more central place in science education. But some last-minute concessions to creationists led to very different accounts of the decision's significance in national and in regional news.

At issue were changes in the wording of the science framework that were negotiated in the past few weeks to guarantee the board's approval. The board vote attracted widespread national attention because California controls about 11 per cent of the country's elementary and junior high school textbook market, and its curriculum guidelines — to which publishers must conform — are considered to be influential in setting national standards. The guidelines also influence high school texts, even though the state does not develop an approved list.

The new framework will take effect from 1992 as part of a curriculum overhaul that California undertakes every seven years. Its implementation follows months of public debate over the place of creationism in the science curriculum, which culminated in a public hearing last October.

After that hearing, some board members insisted on several changes to appease Christian fundamentalists. A sentence was deleted that declared that "there is no scientific dispute that evolution has occurred and continues to occur; this is why evolution is regarded as a scientific fact." And phrases were added to the guidelines stating that "some people reject the theory of evolution purely on the basis of religious faith" and that these "personal beliefs should be respected and not demeaned". These changes were heralded by both the *New York Times* and ABC television news as an important victory for creationists, but the *San Francisco Chronicle* referred to the revised guidelines as "a blueprint for teaching science that emphasizes evolution and relegates creationism to social science classes".

But both sides in the dispute agreed that the last-minute changes were relatively minor. Dan Chernow, chairman of the state's curriculum commission, said that while he was disappointed at the changes, he believed that they are "meaningless". And a spokesman for the California-based Traditional Values Coalition said the group was glad for the modifications but called them "crumbs". The revision leaves largely intact repeated references to evolution's fundamental importance to science. For example, the last six words were added to the statement that "the particular case of 'creation science' or 'scientific creationism' has been thoroughly

studied and rejected by the leading scientific societies as qualifying as a scientific explanation". Left unaltered is the declaration that "evolution is the central organizing theory of biology, and has fundamental importance in other sciences as well. It is no more controversial in scientific circles than gravity or electricity is."

Such statements significantly strengthen previous guidelines, which made it clear that evolution must be taught but offered no hints to teachers or school officials on dealing with its role in the controversy surrounding creation theory. The new document addresses this shortcoming, noting in part that teachers "should make it very clear that from the scientific perspective, evolution, like other scientific topics, does not bear on an individual's religious beliefs. Science is not theistic, nor is it atheistic; it does not presuppose religious explanations." Chernow said creation theory should be discussed along with other religious issues, either in the history and social science curriculum or the English and language arts curriculum.

Robert Buderl

ANTARCTIC ASTRONOMY

Cerenkov telescope for the South Pole

Washington

PHYSICISTS from Purdue University and the University of Wisconsin at Madison will be travelling south next month to begin a three-year project that will see construction of the first optical telescope, albeit an unconventional one, at the South Pole. An array of ten 1-metre mirrors will be set out on the ice to look for flashes of light caused by ultra-high-energy particles shooting through the upper atmosphere.

For a number of years, researchers from the University of Leeds and the Bartol Institute at the University of Delaware have been operating an air-shower detector at the South Pole (see *Nature* 340, 192; 1989). This device registers the cascade of debris created by a high-energy particle hitting an atom in the upper atmosphere. The Purdue-Wisconsin group will use an alternative detection method which searches for the Cerenkov radiation emitted by a charged particle moving through a refractive medium such as air.

According to James Gaidos of Purdue, the first goal of the Antarctic project is to demonstrate that optical-quality mirrors will work successfully in the extreme climate. The prototype mirrors for the Purdue-Wisconsin telescope are backed with heating elements to keep their surfaces a few degrees above the air temperature, which should prevent condensation and fogging.

David Lindley

Research for sale

Sydney

A DISPUTE at Melbourne University over a proposal for commercially sponsored professorial appointments has cast academic freedom and industrial needs as enemies. The university intends to create a new category of position by which private companies would select researchers and finance a university chair for them. The idea follows the recommendation of a report from a working party headed by the university's vice chancellor, David Penington, which says that corporate-funded appointments are necessary because the university has run out of money to create professorial appointments in new fields.

The sponsored professorial fellows would be appointed at the request of the donor company, without being subject to competitive selection. Nevertheless, Penington says that "for each corporate-backed professorial chair the University council must be satisfied that there are no strings attached. If a company wants to donate funds to the university to take on a particular individual, we would be foolish to appoint anyone else. We will, however, make sure of the calibre of the appointee." But the Federation of Australian University Staff Associations (FAUSA) is unhappy with the university's proposed safeguards. FAUSA believes that people appointed at the donor's behest may be



reluctant to participate in research that the donor does not approve.

"If we give up academic independence and criticism for the money these corporations will bring in then the price is too high", says Ralph Hall, president of FAUSA. But Penington disagrees: "We must develop links with industry if research and development in Australia is to flourish."

Tania Ewing

Correction

Dr Gerrit Viljoen, the South African Minister for National Education, was professor of Greek at the University of Pretoria before becoming rector of the Rand Afrikaans University in Johannesburg (*Nature* 342, 8; 1989). □

Freedom to conduct research

SIR—The near coincidence of the publication of letters from Donald Kennedy (president of Stanford University; *Nature* 341, 180; 1989) and Norman R. Saunders of Southampton (*Nature* 341, 99; 1989) prompts us to write. The former was in regard to Alex Comfort's claim that academics generally show cowardice in the face of external attack, while the latter dealt in part with the inappropriate comments of Clive Hollands (*Nature* 339, 248; 1989) about a paper from the INSERM group on development of the primate visual system (*Nature* 337, 265–267; 1989). We were beginning to think that Colin Blakemore (*Nature* 339, 414; 1989) was the only scientist in Britain with the courage to speak out on an issue of fundamental importance to all scientists: namely, the freedom to conduct research without unwarranted obstruction or meddling. The reticence of scientists on both sides of the Atlantic to speak out in response to attacks by antivivisectionists such as Hollands tends to confirm Comfort's assertion.

We, too, question the propriety of a member of Britain's Animal Procedures Committee (APC) passing moral judgement on research he apparently knew little about in such a public forum as the Correspondence pages of *Nature*. At the very least, Hollands' credibility in terms of being seen as fair-minded in carrying out his duties on the APC has been compromised. Further, Hollands' decision to identify himself in the letter as a member of the Committee for the Reform of Animal Experimentation is troubling, and raises the question: is it appropriate for members of animal research 'reform' groups to serve on committees that regulate animal research? Do we really

ECU too

SIR—Fred Hoyle (*Nature* 341, 380; 1989) finds that the term ECU (European Currency Unit) induces pictures of infinities of little grey men in a very large grey building in his mind. If he resents it, he will find the image quite changed.

The word "ecu" sounds to me like something small, hard and shiny (for example money). It is pronounceable in every European language (with the obvious exception of Greek, which does not have a "c"), and it is short enough to require no further abbreviation.

Provided it does not mean anything foul or ridiculous in any European language, why look further?

ANDREW JENKINS

AS Telelab, Strømdaljordet 4,
Postboks 1868, Gulset,
3701 Skien, Norway

want foxes minding our research hen-houses?

We do not question the necessity of regulating animal research, nor do we question the desirability of having as members of regulatory committees people who are truly concerned about animal welfare. Indeed, everyone who works with animals in any way should be concerned with their welfare, and happily the vast majority of scientists are so concerned. Our apprehension lies with empowering 'false moderates' who cloak their true antivivisection/antiscience agenda with euphemisms. Is this paranoia? We think not. As you point out in your leading article (*Nature* 339, 491; 1989), conditions for and treatment of animals in research laboratories are improving, in part because of the clamour made by animal activists, many of whom claim to be concerned only with improving welfare. But the hypocrisy of this claim is well illustrated by recent experiences of our university and nearby Stanford. Both universities have been making great efforts at considerable cost to replace antiquated research facilities with state-of-the-art laboratories that will greatly improve the animals' environment. Nevertheless, these changes have been met by the animal 'welfare' advocates not with appreciation but with lawsuits and ever-increasing pressure to halt construction. It seems that every new regulation, stipulation or increased cost of doing animal research is viewed by the antivivisectionists as a victory, and causes them to redouble their efforts to halt such research by whatever means they choose to use. For the 'moderates', it means more lobbying for more expensive and restrictive regulations; for the extremists it means vandalism, theft, arson, bombs and death threats (so far). The silence of the vast majority of scientists on this issue may ensure success for the antivivisectionists, and once again, in the words of Dr Mary Putnam Jacobi when she testified in 1900 against a bill that would have curtailed animal research in the District of Columbia, we shall see "the domination of knowledge by ignorance".

CHARLES S. NICOLL
SHARON M. RUSSELL

Department of Integrative Biology,
University of California, Berkeley,
Berkeley, California 94720, USA

SIR—The US Department of Agriculture (USDA)'s "dramatic about-turn" allowing research institutions "themselves to decide how to meet the required standards" of animal care (*Nature* 341, 6; 1989) will add to the already substantial abuse of animals in laboratories. As it is, USDA officials rarely inspect laboratories

more than once a year, if that. Violations of the Animal Welfare Act often go undetected, and if they are reported, there may be no follow-up to ensure compliance with the law. Animal protection organizations have exposed blatant violations and abuse that would otherwise still occur.

USDA's authority over laboratories needs to be strengthened, as does the Animal Welfare Act. Allowing researchers to police themselves will result in conditions that benefit the research budget, not the animals.

K. S. GUILLERMO

People for the Ethical
Treatment of Animals,
PO Box 42516,
Washington, DC 20015, USA

Nature conservancy

SIR—As the British government's proposals to split up the Nature Conservancy Council (NCC) will undoubtedly generate much public debate during the next 18 months or so, I should like to correct some inaccuracies in Ben Webb's recent article (*Nature* 341, 94; 1989).

First, NCC has no statutory responsibility for archaeological sites. Its remit covers the conservation of flora and fauna as well as features of geological or physiological importance.

Second, before 1973, the then Nature Conservancy did carry out research, but following the formation of NCC this function was handed over, in part only, to the Institute of Terrestrial Ecology (ITE). But NCC's Chief Scientist's Directorate (CSD) coordinates or conducts a wide range of research and survey work necessary for NCC to exercise its diverse functions throughout Great Britain. To say that those opposing the government's proposed reorganization are unlikely to "seek to put the clock back to when NCC had a research capability of its own" is misleading, NCC utilizes such a research capability today. Indeed, the CSD provides the science base and overview upon which many areas of the NCC's work are justified and built.

Third, while Webb focuses on the need for plans to encourage farmers to take into account environmental considerations, no mention is made of other land uses and development pressures. The forthcoming debate will by no means linger on the farming community; many of the problems of conservation in Great Britain today are caused by other socio-economic factors related to industrial and residential development, transport and infrastructure provisioning, forestry and recreational pressures on seminatural and other habitats of importance for wildlife.

PETER J. EWINS

Nature Conservancy Council,
Archway House, 7 Eastcheap,
Letchworth, Herts SG6 3DG, UK

Edging towards human gene therapy

False starts in identifying genes associated with psychiatric disease (see overleaf) and excessive regulatory hurdles facing human gene therapy could hamper research on gene manipulation. But the mood was optimistic at *Nature's* conference on Gene Manipulation in Biology and Human Disease.

Boston

INTRODUCING a session on the prospects for gene therapy at *Nature's* conference in Boston last week, Leon Rosenberg (Yale University) wisely refrained from predicting when the first trials of human gene therapy would commence. Progress towards them, he said, has been perhaps slower than some had hoped but faster than others had feared. The fact of the matter is that potential gene therapists have been beaver away to produce appropriate reagents and procedures in the full knowledge that, for no good reason, they will face far more barriers than are usual for experimental clinical trials.

Among the more genuine reasons why any trial of gene therapy is still some years off, is that many of the technical problems have been hard to crack. Another is the continuing memory of a premature attempt in the early 1980s to correct the faulty β -globin gene in two cases of thalassaemia; the attempts were mounted in Israel and Italy, but were led by a US researcher, whose research efforts have since been redirected.

It is ill wind, however, that blows no good. Those interested in globin gene therapy were forced to take stock. First, it became clear that the control of globin gene expression is more complex than had been thought. The frustrating result was that whereas it was possible to insert globin genes into cells, insufficient globin was produced to make the cells of any potential use in therapy. But a solution to that problem now seems to be within reach: DNA sequences that confer high expression on globin genes, despite being located far from them, have been identified and shown still to function after being engineered to be much closer to the genes.

In the meantime, ironically, the need for gene therapy in thalassaemia has receded, as bone marrow transplantation has increasingly found a place in the treatment of the most serious forms of the disease. Used in that way, transplantation is a type of gene therapy, albeit one that cocks a snook at the sophistication of genetic engineering. Whereas its use is limited by the availability of compatible bone marrow and the problems of graft rejection, both problems could, in principle, be solved were it possible to transplant the stem cells that are the progenitors for all

blood cells rather than whole bone marrow.

Such a prospect was held out by Irving Weissman (Stanford University). As the much-needed assay for human stem cells, he hopes to test their ability to reconstitute the haematopoietic system of the genetically immunodeficient SCID mouse. But, to judge by his experience with mouse stem cells, there would still remain the pressing problem of finding a practical way to stimulate the proliferation of isolated human stem cells if they are ever to be available in sufficient quantities for transplantation.

In principle, many genetic deficiencies could be treated by transplantation of the appropriate normal tissue, but because of the practical difficulties surrogate methods are being explored. One example that excited interest at the meeting was the 'neo-organs' or 'organoid neovascular structures' approach described by W. French Anderson (National Institutes of Health). Neo-organs are the end result of implanting rats with fibres of polytetrafluoroethylene (Gortex) coated with collagen, heparin binding growth factor 1 (also known as acidic fibroblast growth factor) and hepatocytes. After several months, the implanted structures have acquired a continuous blood supply from the liver, contain structures that resemble nerves and have grown to as much as half the size of the liver. That hepatocytes in neo-organs are functional has been shown by the apparent ability of neo-organs seeded with Lewis rat hepatocytes to restore to normal, for at least six months, the elevated levels of serum bilirubin in Gunn rats (Gunn rats have a genetic deficiency of the enzyme UDP-glucuronosyltransferase).

Even if cell or organ transplants may limit the need for specific gene therapy in some cases, the ingenuity of gene manipulators ensures that they will continue to thrive. Not all will be lost, for example, should there be no need to use β -globin gene therapy to treat thalassaemia, because knowledge of how to obtain adequate globin gene expression is already being used to develop a mouse model of sickle-cell anaemia in which to test new forms of therapy. The first step, as described by Richard Palmiter (University of Washington, Seattle), is to produce transgenic mice carrying the mouse equivalent of the mutant α -globin gene

that is the cause of the human disease. When these mice are crossed with a strain that lacks one copy of the β -globin gene, the resulting offspring have red blood cells that will sickle.

Mice carrying foreign genes, whether introduced into fertilized eggs (transgenic technology) or embryonic stem cells, are proving to be rich sources of animal models of human diseases. For example, neuroblastoma can be modelled by mice carrying a polyoma virus gene (Erwin Wagner, Institute of Molecular Pathology, Vienna), and the liver damage progressing to malignancy associated with hepatitis B virus infection can be modelled by a hybrid transgene composed of a hepatitis viral gene and an albumin gene promoter (Palmiter). A variation on this theme is to try and interfere with the function of a gene in order to produce a model of a genetic deficiency. Thus the cleft palates produced in transgenic mice carrying a hybrid viral *erbA*/retinoic acid receptor gene, provide a model in which to explore the possible role of retinoic acid in morphological development (Ronald Evans, Salk Institute).

The transgene in that case is a so-called dominant negative. It is the possibility of expressing in the blood cells of AIDS patients dominant negative genes for proteins of the human immunodeficiency virus (David Baltimore, Whitehead Institute) in order to interfere with human immunodeficiency virus function, that might yet drive the first successful attack on the growing battery of regulations that face anyone wishing to pioneer the use of gene therapy.

For now, the only tentative step towards such therapy that has been taken is the inclusion of a genetic marker in some of the 'tumour infiltrating lymphocytes' that are being used in an experimental procedure for cancer treatment. The marker is enabling the fate of the cells to be followed (Anderson) but plans are already afoot to include not just a marker gene but also a gene for interleukin-2 that may lengthen the life and efficacy of the cells. As their use will be tested only in patients for whom conventional forms of treatment have failed, there should be no more reason for excessive regulatory hurdles than is usual in such circumstances. Unfortunately, given that gene therapy is involved, even that is too much to hope for.

Peter Newmark

False start on manic depression

Miranda Robertson

MOST of the presentations at *Nature's* conference in Boston, held last week, were a startling reflection of the pace of progress in the understanding of the genetic mechanisms of complex processes. The one conspicuous reversal, candidly acknowledged by Ken Kidd (Yale University), was in the genetics of psychiatric disease.

Readers of *The New York Times* may already be aware that the linkage reported in *Nature* two years ago¹ between manic depression and a locus on chromosome 11 is now acknowledged probably to have been spurious. The events that led to this conclusion are described, by a group including many of the authors of the original report, on page 238 of this issue², and were discussed at the meeting by Kidd who is co-author on both papers. At the end of his presentation, David Baltimore (Whitehead Institute) voiced the question that must have been left in many people's minds: "Setting myself up as an average reader of *Nature*", he asked, "what am I to believe?". What follows is an attempt to enable readers to answer the question for themselves.

A picturesque feature of the original study was the choice of the Amish community of Pennsylvania for analysis. The Amish are no more prone to psychiatric disease than other groups, but they have three advantages as subjects for the investigation of complex genetic traits. First, they have large families whose members do not disperse. Second, the restrictive rules they live by reduce confounding environmental variables such as drugs and drink. Third (and crucially), the Amish are relatively inbred, which increases the probability that an abnormal phenotype running in a family will be traceable to the same gene in all affected members.

The claim of the 1987 paper was that manic depression in a large Amish pedigree was linked to two marker genes, the *Harvey ras* oncogene and the insulin gene, known to be in the same region of the short arm of chromosome 11. Neither of these genes was itself implicated in the disease, nor was any other gene actually identified as the cause of the disorder: the claim rested upon the consistent inheritance of specific and recognizable variants of the two marker genes with the propensity to develop manic depression, and on a calculation of the probability that this co-inheritance could occur purely by chance. The traditional measure of this probability is the LOD score (LOD standing for the log-of-the-odds ratio), and the traditionally acceptable LOD value is 3, corresponding to a probability of 0.05 against the chance occurrence of the observed

pattern of co-inheritance. The LOD score for the linkage of a chromosome 11 locus to manic depression was comfortably more than 4.

For any disorder, the assessment of linkage is complicated by the possibility of recombination events that may separate the marker from the gene and that become more likely the greater the distance between the two (this distance being of course unknown). For psychiatric disorders the calculation is further complicated by the likelihood of incomplete penetrance, which means that some individuals with the predisposing gene will fail to develop the disorder. Both factors must be taken into account in calculating linkages to psychiatric diseases.

Thus the conclusion that a gene on chromosome 11 could predispose to manic depression was calculated on a delicate balance of uncertainties. Two developments have now upset that balance. First, two people in the original Amish pedigree have become manic depressive in the absence of the putative markers of predisposition; second, an additional branch of the original pedigree shows negative evidence of linkage. This could mean that there are two genes instead of one predisposing to manic depression in the pedigree; but the simplest explanation is that the apparent linkage to a chromosome 11 locus was in fact just chance.

According to Kidd, chance may also account for a second genetic linkage, more recently reported by Hugh Gurling's group³ which has published evidence implicating a locus on chromosome 5 in predisposition to schizophrenia. The conclusion in that case was based on two markers which, as Lander⁴ remarked at the time, were uninformative for many members of the five families analysed, where the same marker variants were carried by both the affected and the unaffected parent.

In an attempt to find a marker that will distinguish the parental chromosomes in these cases, Gurling has since tested a third chromosome 5 marker which, as it turns out, shows less evidence of linkage than either of the other two. This could mean that the third marker is more distant from the putative schizophrenia locus than the others, allowing more recombination between the marker and the gene; but in fact Kidd maps the third marker in between the other two, ruling out that explanation. This leaves us with no persuasive evidence linking any psychiatric disease to a single locus and brings us back to Baltimore's question.

It is neither necessary nor justified to conclude from these reversals that there

simply is no single gene capable of predisposing some individuals to psychiatric disease. Studies of identical twins have already provided very strong evidence, especially in the case of schizophrenia, for an important genetic component. There is on the other hand every reason to expect the predisposing genes to be hard to identify, for reasons that can be illustrated by the remarkably successful recent attempts at analysing the complex multifactorial genetics of predisposition to coronary heart disease.

The work on cholesterol metabolism in particular was discussed by Gerd Utermann (University of Innsbruck), who has been investigating the effects of plasma lipoprotein polymorphisms on coronary heart disease in people already known to be at risk through a mutation in one of their low-density lipoprotein (LDL) receptor genes. People who are homozygous for such mutations never escape heart attack. Heterozygotes on the other hand may have heart attacks by the age of 40, or they may be quite unaffected. What Utermann has been able to show is that polymorphisms in plasma proteins that affect cholesterol levels can account for much of the variation in effect of the LDL receptor mutation on heterozygotes.

In this case several interacting genes affecting the uptake and disposition of lipids can be seen to contribute to the risk of heart disease in such a way that no one mutant gene or polymorphic variant would reliably be inherited with the disease. We may expect the genetics of psychiatric disease to be at least as complex as those of cholesterol metabolism. But in the absence of a good biochemical clue to what genes may be involved, they will be much harder to trace. The hope has been that the high incidence of specific psychiatric disorders in inbred families reflects the existence of a single, strong, predisposing allele. If in fact it is necessary to seek linkage to more than one gene in such families, there are techniques that would make this possible⁵; but a high-resolution linkage map of polymorphic markers for the human genome would be needed for a reasonable chance of success, and in the view of many this should be a priority for the human genome project.

In the meantime, readers of *Nature* are entitled to be sceptical about reports of psychiatric disease linkages at LOD scores of less than 6. But they should have no reason to doubt the existence of genetic predisposition to psychiatric disease, nor the ability of molecular geneticists eventually to identify the genes responsible. □

Miranda Robertson is Biology Editor of *Nature*.

1. Egeland, J. A. et al. *Nature* **325**, 783-787 (1987).
2. Kelsoe, J. R. et al. *Nature* **342**, 238-242 (1989).
3. Sherrington, R. et al. *Nature* **336**, 164-170 (1988).
4. Lander, E. S. *Nature* **336**, 105-106 (1988).
5. Lander, E. S. & Botstein, D. *Cold Spring Harb. Symp. quant. Biol.* **49**, 49-62 (1986).

Microwaves in random media

J. B. Pendry

A DIFFICULT statistical problem, as yet unsolved, concerns the motion of waves in random media. The problem can be compared with the kinetic theory of gases, one of the great achievements of physics. Newton's equations of motion were redeveloped to formulate equations describing the macroscopic properties of the gas. Crucial to the process was the interplay between theoretical insight and experimental stimulation. In the case of waves in random media experimental results will be equally critical. A recent experiment by N. Garcia and A. Z. Genack (*Phys. Rev. Lett.* **63**, 1678–1681; 1989) provides a veritable paradigm of the problem for theorists.

It is no accident that the kinetic theory of gases was developed long before theorists felt able to tackle waves in random media. A large sample of gas can be characterized by macroscopic quantities such as temperature, pressure and volume, which are well defined and whose statistical fluctuations can be neglected. Not so for waves in a random medium: the further waves penetrate into a disordered system the more extreme become the intensity fluctuations. This is true for any degree of disorder; but for strong disorder, the wave is essentially scattered within its own wavelength and the resulting wavefield is even more statistically complex. The fluctuations cannot be ignored and must be included in any theory.

An example of these fluctuations is given in Garcia and Genack's article. The sample consists of uniform half-inch-diameter polystyrene spheres with filling fraction 0.56, contained in a 7.3-cm-diameter copper tube. The wavelength of microwave radiation used was about 1 cm. The spheres fill the tube uniformly,

but can still move when the tube is tumbled, providing another statistical configuration. Microwaves are launched from a horn 20 cm in front of the sample and detected by a Schottky diode mounted on a plunger that defines the length through which the microwaves pass. Figure 1a shows the transmitted intensity as a function of frequency for a static sample 140 cm in length. Note the extreme fluctuations, which take the form of a series of sharp peaks separated by deep troughs. Tumbling the tube generates a different set of peaks and troughs, but of the same general character.

If these results were unique to microwaves the experiment would be a mere curiosity. Let us take a couple of other examples from the pages of *Nature*. A few months ago in News and Views (*Nature* **339**, 581; 1989), I discussed waves in semiconductors: in that instance the waves were, of course, electrons. When the semiconductor is disordered, the conductance–voltage plot may look like that shown in Fig. 1b. These data (A. B. Fowler, J. J. Wainer & R. A. Webb *IBM J. Res. Devel.* **32**, 372–383; 1988) are for a silicon–metal–oxide field-effect transistor at 50 mK. The similarity in form to the microwave transmission experiments is striking because the origin of the fluctuations is the same: waves in random media.

An earlier example comes from W. H. Munk and G. O. Williams (*Nature* **267**, 774–778; 1977). This paper discussed the transmission of 406 Hz acoustic signals through the ocean. The distances in question were very long — of the order of 1,000 km — and the disorder results from mesoscale fluctuations in the ocean's currents. A simulation of the acoustic transmission through a typical path, plotted in

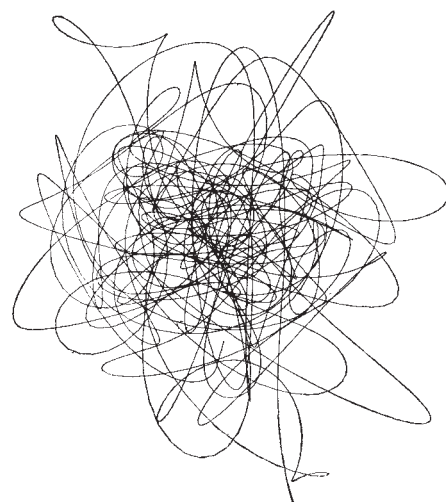


FIG. 2 Acoustic transmission simulated for transmission through approximately 1,000 km of fluctuating ocean at a frequency of 406 Hz. The distance from the origin shows the intensity; the angle, the phase.

the complex plane (Fig. 2), reveals both amplitude and phase fluctuations. It is clear that extreme fluctuations are again involved. A related phenomenon is the fading of radio waves in the ionosphere. An example closer to common experience is provided by the propagation of a thunder clap across a random landscape: close to the strike a single sharp crack is heard; farther away the Fourier components in the crack pass through the random acoustic filter provided by a disordered landscape, and recombine to produce the characteristic rumble whose long duration bears witness to the narrow bandwidth of the transmission peaks.

Of these examples, the electronic instance is the most important, with its consequences for problems likely to be encountered in the construction of small low-dimensional devices. Despite the rich variety of experiments on disordered electronic systems, there are some problems involved in comparing them with the current imperfectly developed theories. First, the scale of electronic wavelengths is such that probes on this scale have yet to be developed. Second, electrons are mutually interacting particles and the complication of treating disorder at the same time as electron–electron interactions is one that the theorist would rather be without until a clear understanding emerges of the effects of disorder alone. Third, electronic experiments are, as a rule, made on a single device. Switching to another device is often time consuming and it is impossible to perform ensemble averaging over specimens. This can be a great disadvantage where strong fluctuations are important.

The microwave experiment eliminates all these objections and should prove an ideal test bed to knock on the head unsuccessful theories, and to help good ones.

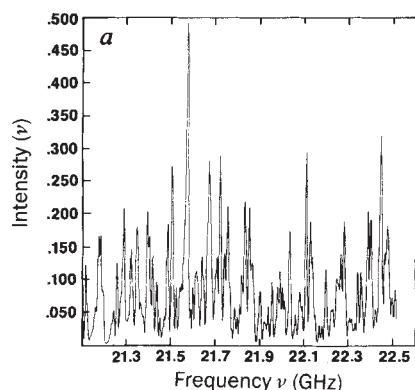
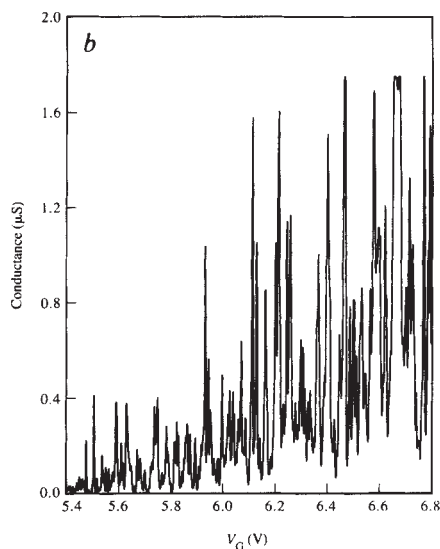


FIG. 1a, Spectral intensity fluctuations for microwaves passing through a 140-cm length of tube filled with half-inch-diameter polystyrene balls. b, Conductance of a silicon–metal–oxide field-effect transistor as a function of gate voltage measured at 50 mK.



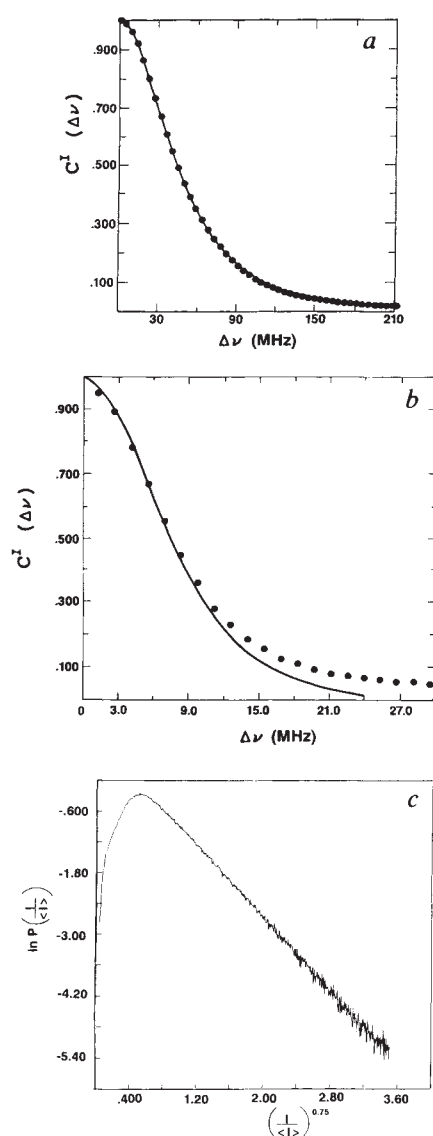


FIG. 3 The correlation function $\langle I(\nu)I(\nu + \Delta\nu) \rangle$ for transmitted intensity at two frequencies separated by $\Delta\nu$. The solid line shows the theoretical fit. *a*, Length of tube is 30 cm; *b*, length of tube is 140 cm. In the longer sample, strong-disorder effects begin to assert themselves and the correlation function shows a high-frequency tail not explained by the theory. *c*, The probability distribution of intensities transmitted through a tube of length 140 cm, plotted against $(I/I_0)^{3/4}$.

Figure 3 provides an instance of this process at work. It shows intensity-intensity correlation functions from Garcia and Genack's paper where the intensities are measured with frequency separation $\Delta\nu$. Theory developed for the weak-scattering limit fits the data very well for short samples (Fig. 3*a*), but not for longer ones in which the scattering builds up to large values (Fig. 3*b*). Another puzzle is shown in Fig. 3*c*, a plot of the probability distribution of transmitted intensities. A curious exponent of 3/4 is found for the dependence of $\log(P(I))$ on I . So far theory does not have an explanation.

Theorists are not completely at sea

with these problems. Even for strongly disordered samples we have a general understanding of why strong fluctuations occur. Disorder strongly inhibits the passage of waves through the medium, and therefore it is possible for resonant cavities to form at random within the medium. Waves can be trapped in these for a long time before tunnelling out. The longer they are trapped, the sharper the resonances, as we understand from conventional theory of resonances, so that a thick, strongly disordered sample will have more sharp resonances than a short, weakly disordered one. The resonances play an important role in the transmission properties of the sample: waves whose frequency is degenerate with a resonance show enhanced transmission. The sharp spikes seen in Fig. 1*a* betray the presence of resonant cavities within the sample.

PROTEIN STRUCTURE

Chaperones, paperones

Joe Sambrook and Mary-Jane Gething

FOR the past several years, chaperone proteins have been a valuable commodity for speculators in protein function. Identified in most organelles of eukaryotic cells, chaperones bind to newly synthesized polypeptides and apparently stabilize them until they are assembled into their correct native structure^{1,2} or until they are transported into another cellular compartment. With the description by Holmgren and Brändén on page 248 of this issue³ of the first three-dimensional structure of a chaperone protein, albeit one that is bacterial and rather unusual in several respects, comes a result that was not predicted by any of the speculators: the two structural domains of the protein that seem to carry out the chaperone function resemble immunoglobulin domains.

Most of our knowledge about chaperones has come from biochemical and genetic studies carried out in eukaryotic cells^{3,5-8}. As far as we know from these studies, chaperones are not true catalysts of protein folding. They are, for example, unable to catalyse the formation or isomerization of covalent bonds in the target proteins. In these respects, chaperones differ from conventional enzymes such as protein disulphide isomerase, prolyl isomerase or prolyl hydroxylase, that modify the structure of newly synthesized proteins. In addition, chaperones can remain attached to their target proteins for far longer than an enzyme would be expected to interact with its substrate⁹. Finally, many chaperones can interact with a wide variety of target proteins *in vivo*^{1,2} and with synthetic peptides of diverse sequence *in vitro*¹⁰. These targets seem not to have any amino-acid sequences or structural motifs in common.

The statistics of transmission through strongly disordered samples is the statistics of these resonances.

A whole series of new experiments is opened up to us. What are the effects of exchanging the one-dimensional tube with a two-dimensional sheet? Can the reflection coefficient be measured? And how do the resonances beneath the surface affect the noise characteristics of the reflected signal? Can a more complex medium be studied, for example one in which the dielectric medium has lower symmetry and shows optical activity? What about the higher moments of the transmission coefficient? The answers to these questions are eagerly awaited. □

J. B. Pendry is in the Blackett Laboratory, Imperial College of Science, Technology and Medicine, London SW7 2BZ, UK.

Diversity on such a scale is not found in substrates for conventional enzymes.

Most of the experimental evidence is, instead, compatible with the idea that chaperones stabilize segments of polypeptide that might otherwise serve as local centres for intra- or interchain aggregation or denaturation. In this way, chaperones might provide the time needed for thermodynamic forces to drive folding of large and complex proteins to completion. According to this idea, chaperones would increase the efficiency of folding and assembly of newly synthesized proteins, but would not necessarily expand the repertoire of interactions that occur between polypeptide chains in the final protein. Alternatively, they might also maintain parts of the nascent polypeptide chain in configurations that could eventually be converted into a wider variety of shapes and surface than would otherwise be possible. Chaperones would therefore not only increase the efficiency of protein folding but would also extend the range of interactions both within and between polypeptides. A simple analogy is the lock and key structures that hold together poppet heads in a child's necklace. Although such interlocking structures are unknown in proteins, there are many examples of interactions between deeply convex and concave surfaces. Chaperones may serve as a mechanical jig to hold segments of a polypeptide chain in a state that allows complex surfaces to form. Perhaps, without chaperones, only a subset of complex topological surfaces could be generated by interaction between 'pre-assembled' polypeptides.

Holmgren and Brändén's crystal structure, which provides a new window on

chaperone mechanisms, is of the PapD protein isolated from the periplasmic space of *Escherichia coli*. Strains of *E. coli* that cause nephritis carry special pili (aptly named P pili) that allow the bacteria to adhere to digalactoside residues on the surface of cells lining the urinary tract^{11,12}. These pili, which project through the outer bacterial membrane and cell wall, consist of approximately 1,000 protein subunits and are assembled in the periplasmic space from components that have been translocated across the inner bacterial membrane. The shaft of each pilus consists of repeating units of a single protein (PapA), whereas the tip consists of three proteins, PapE, PapF and PapG¹³, the last of which carries the digalactoside-binding activity of the pilus¹⁴.

The single operon that encodes all of these proteins also encodes two accessory proteins that are not incorporated into the pilus but are required for its formation^{12,15}. One of them, PapC, appears to form a channel in the outer bacterial membrane through which the growing pilus emerges¹⁶. The other, PapD, is a periplasmic protein that forms complexes with subunits of at least two of the proteins (PapG and PapE) before they are assembled into the pilus¹⁷. Earlier this year, a group of Swedish workers proposed that PapD acts as a chaperone to enhance the postranslational folding and assembly of components of the pilus¹⁸.

Holmgren and Bränden now show that PapD is built from two globular domains, each consisting of an antiparallel structure formed from packed β -sheets. The topology of each domain is the same as that of the constant domain of the immunoglobulins. Whereas there is no significant amino-acid similarity between either of the two domains of PapD and the constant domain of immunoglobulins, the sequence of 239 amino-acid residues can be aligned with the N-terminal 256 amino-acid residues of the Leu-1 human lymphocyte differentiation antigen (also known as CD5). Of the amino-acid residues in the aligned sequences, 26 per cent are identical. The authors propose that PapD and CD5 form a special subclass of the immunoglobulin gene superfamily. Whether or not their speculation that the subclass was dispersed by horizontal gene transfer during evolution turns out to be true, their work provides dramatic evidence of the durability and utility of this type of protein fold.

Chaperone watchers will be most interested in the region of PapD that potentially interacts with its target proteins. The best candidate is the wide crevice between the two domains that contains an exposed hydrophobic patch flanked by basic and acidic residues. In addition, the region linking the two domains perhaps serves as a flexible hinge that can open up to

accommodate target sequences of various sizes. The orientation of the target polypeptide within this site is not known, as the PapD molecules crystallized by Holmgren and Bränden are 'empty'. It would be interesting to find out whether PapD, like chaperones such as BiP and HSP70 (ref. 10), can bind oligopeptides of defined sequence and, if so, whether the binding is tight enough to allow crystallographic analysis of PapD-target complexes.

PapD is an unusual chaperone in several respects. It is smaller than most, contains no obvious ATP-binding sites and appears to be more specific in its function. Unlike BiP or other members of the HSP 70 family, which bind to a wide variety of target proteins and perform functions that are indispensable to the life of the cell, PapD binds only to polypeptides encoded by the *pap* operon. Therefore it is not necessary for the survival of *E. coli*, but only for its colonization of the urinary tract. Because the function of PapD is more specialized than that of other chaperones, the potential polypeptide-chain binding site formed between two immunoglobulin-like domains may differ in detail from those of chaperone proteins that fulfil more generalized functions in protein folding and assembly. There cannot, however, be too many ways to

form clefts that can accommodate polypeptide segments of varying sequences. Holmgren and Bränden's structure, together with that of the class I major histocompatibility antigen^{19,20}, provide examples of two ways to solve this problem. It remains to be seen whether other chaperones use similar structural motifs. □

Joe Sambrook and Mary-Jane Gething are at the Southwestern Medical School, Dallas, Texas 75235, USA.

1. Ellis, J. *Nature* **328**, 378–379 (1987).
2. Ellis, R.J. *Trends biochem. Sci.* **14**, 339–342 (1989).
3. Deshaies, R. & Schekman, R. *Trends biochem. Sci.* **13**, 384–388 (1988).
4. Holmgren, A. & Bränden, C.-I. *Nature* **342**, 248–251 (1989).
5. Pelham, H. *Cell* **46**, 959–961 (1986).
6. Deshaies, R.J. *et al.* *Nature* **332**, 800–805 (1988).
7. Normington, K. *et al.* *Cell* **57**, 1211–1236 (1989).
8. Rose, M.D. *et al.* *Cell* **57**, 1211–1221 (1989).
9. Bole, D.G. *et al.* *J. Cell. Biol.* **102**, 1558–1566 (1986).
10. Flynn, G.C. *et al.* *Science* **245**, 385–390 (1989).
11. Norgren, M. *et al.* *EMBO J.* **3**, 1159–1165 (1984).
12. Normark, S. *et al.* in *Microbial Lectins and Agglutinins: Properties and Biological Activity*. (ed Mirelman, D.) 113–143 (Wiley Interscience, New York, 1986).
13. Lindberg, F. *et al.* *Nature* **328**, 84–87 (1987).
14. Lund, B. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 5898–5902 (1987).
15. Lindberg, F.P. *et al.* *EMBO J.* **3**, 1173–1176 (1984).
16. Norgren, M. *et al.* *Mol. Microbiol.* **1**, 169–178 (1987).
17. Lindberg, F.P. thesis, Umeå Univ. (1987).
18. Hultgren, S.J. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **86**, 4357–4361 (1989).
19. Bjorkman, P. *et al.* *Nature* **329**, 512–518 (1987).
20. Bjorkman, P. *et al.* *Nature* **329**, 506–512 (1987).

OZONE DEPLETION

A dent outside the hole?

Guy P. Brasseur

THE formation of an ozone hole inside the polar vortex over the Antarctic continent each spring (September to November) since the late 1970s has been a cause of wide concern. In a rather provocative paper on page 233 of this issue¹, Proffitt *et al.* suggest that, from August to September, a substantial amount of high-latitude ozone is also destroyed outside the Antarctic vortex, by unidentified chemical processes. This conclusion is reached by an indirect but elegant analysis of simultaneous measurements of ozone, nitrous oxide (N_2O) and reactive nitrogen species (abbreviated as NO_x) made by stratospheric jet-borne experiments.

Following the discovery of the ozone hole, there were several coordinated measurement expeditions, including the 1987 Antarctic Airborne Ozone Experiment from which Proffitt *et al.* derive their data, in an attempt to elucidate the mechanisms responsible for the depletion. These have revealed that active ozone-destroying forms of chlorine are produced in the lower stratosphere by the conversion of relatively stable molecules such as HCl and $ClONO_2$ on the surface of ice particles in the stratospheric clouds formed in the polar winter (when temperatures drop below $-85^\circ C$). The

concentration of chlorine monoxide inside the polar vortex during the Antarctic spring is found to be 100 times higher than at midlatitudes: this is sufficiently elevated to destroy much of the ozone in a couple of weeks.

The suggestion by Proffitt *et al.*¹ of early-springtime depletion outside the polar vortex is a surprise because no such mechanism can operate there. The first step in the analysis leading to their suggestion is to note that, outside the chemically perturbed region of Antarctica, there is a persistent linear (negative) relation between the concentration of N_2O and NO_x in the lower stratosphere. The source of nitrogen oxides in the stratosphere is the slowly reactive nitrous oxide, which is produced at the Earth's surface and transported into the stratosphere by air currents. The necessary conservation of nitrogen atoms inside a given air parcel, as the conversion from N_2O to NO_x proceeds, explains the relation between the concentration of these species. The relation, however, does not hold in chemically perturbed regions, where NO_x is efficiently removed by ice particles and by their subsequent sedimentation. Thus, by making use of this relation between N_2O and NO_x , it becomes possible to

discriminate between air masses that have been affected by the perturbed chemistry in the region of the ozone hole and those that have not been perturbed.

In the second step of their analysis, Proffitt *et al.* calculate the concentration ratio of O_3 to NO_x along a flight path from California to Chile. If, in the lower stratosphere outside the polar vortex, both ozone and NO_x are not significantly affected by chemistry, this concentration ratio should remain fairly constant. In air masses unperturbed by polar chemistry, this concentration ratio is of the order of, or larger than, 300 at heights of 18–21 km at all latitudes, except south of 50° S where it is substantially reduced. The authors conclude that a significant chemical sink must exist in the region between this latitude and the boundary of the chemically perturbed region, located near 65° S. Losses are about 15 per cent in late August and as high as 30 per cent in late September. These losses were not isolated events but were found throughout the period during which observations were made.

These results are consistent with statistical analyses² of ozone data obtained in this region. In addition, comparisons between vertically integrated ozone concentrations measured between 1979 and 1988 by the total ozone mapping spectrometer (TOMS) on board the Nimbus 7 satellite also indicate substantial ozone reductions in the Southern Hemisphere outside the chemically perturbed region^{3,4}. But, in this case, the depletion was attributed to the dilution in late spring of ozone-poor air originating in the region of the ozone hole. The results presented by Proffitt *et al.* indicate that, even before the polar vortex breaks up, and probably as early as August (before the formation of the ozone hole), an efficient chemical mechanism destroys ozone at high latitudes, south of 50° S. The possibility that ozone is destroyed during the Antarctic winter (as early as June) in the layer where polar stratospheric clouds are present, was recently suggested by Fiocco *et al.*⁵.

Despite all this evidence, no definite conclusion can yet be drawn. The number of ground-based Dobson instruments, by which ozone is measured (and which are used to correct some of the satellite measurements), is too small for the data analyses to be sufficiently reliable. Furthermore, the quantitative values reported by Proffitt *et al.* are somewhat marginal and do not show unambiguously that the anomaly in the O_3/NO_x ratio is a manifestation of chemical destruction. The increased value of this ratio they observed over the tropics is not entirely explained.

But alternative explanations for the measurements by Proffitt *et al.* are readily discounted. For example, the vertical exchange of air parcels outside the vortex

would alter the local composition. As suggested by the high temperatures and maximum ozone (and NO_x) concentrations observed near 60° S, air parcels should subside in this region. But the O_3/NO_x ratio increases slightly with altitude, so that exchange would give the opposite effect to the one reported. Alternatively, enhanced production of NO_x , which is caused by galactic cosmic rays with maximum efficiency at 15–20 km, would result in a reduced ratio of O_3/NO_x . But this process is confined to a region near the magnetic pole that does not extend out to 50° S.

A chemical process thus seems to affect ozone at high latitudes outside the region where the Antarctic ozone hole appears and before the hole is produced. Further evidence for such an effect as well as an explanation of the processes involved needs, however, to be presented before it can be claimed that a significant chemical ozone destruction is taking place in winter and spring. The concentration of active chlorine outside the vortex is not large

enough to account for a 15–30 per cent ozone loss even with the sunlight available at these latitudes during winter, so it seems unlikely that the hole-type loss mechanisms account for the observed ozone destruction. Perhaps heterogeneous chemical processes on the surface of background sulphuric acid aerosols are responsible. Other possibilities are mentioned by Proffitt *et al.* In any case, it will be interesting to see if similar anomalies in the O_3/NO_x concentration ratio are found in the high-latitude region of the Northern Hemisphere. □

Guy P. Brasseur is at the National Center for Atmospheric Research, PO Box 3000, Boulder, Colorado 80307, USA.

1. Proffitt, M.H., Fahey, D.W., Kelly, K.K. & Tuck, A.F. *Nature* **341**, 233–237 (1989).
2. Albritton, D.L. & Watson, R.T. (eds) *Stratospheric Assessment of Stratospheric Ozone* (UNEP/WMO Report, in the press).
3. Atkinson, R.J., Matthews, W.A., Newman, P.A. & Plumb, R.A. *Nature* **340**, 290–294 (1989).
4. Lloyd, P. *Nature* **340**, 269 (1989).
5. Fiocco, G., Komhyr, W.D. & Fuà, D. *Nature* **341**, 346–347 (1989).

PALAEOENVIRONMENTS

No smoke without fire

Peter D. Moore

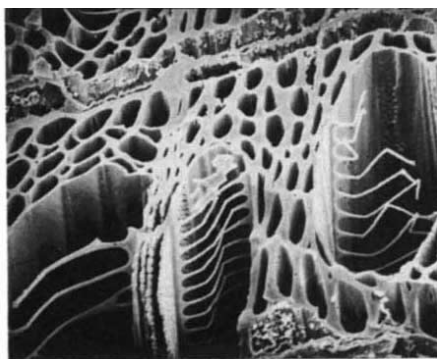
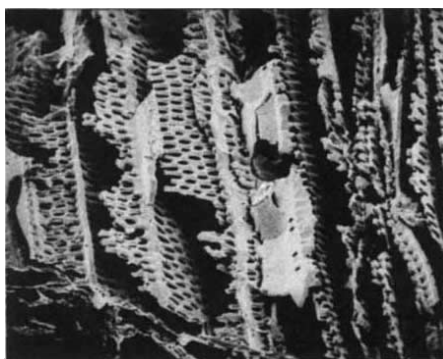
AMONG her many other contributions to science, Marie Stopes, in 1919, analysed the constituents of coal and found particles of black, opaque matter that was chemically inert yet physically fragile. Stopes believed that this material, which she called 'fusain', represented fragments of plant charcoal derived from ancient fires. Fusain, which has subsequently proved to be a common component of coals and many other rocks, has attracted considerable attention in recent years as it could provide evidence of past wildfire frequency and hence of the gaseous and particulate composition of the palaeo-atmosphere. Using scanning electron microscopy and reflectance microscopy to examine charcoal generated from contemporary plant material, Scott¹ has now laid the basis for further understanding the origin and significance of fusain.

The source of ancient fusain has been much debated. Some workers considered it to be a product of wet decay, but Harris² found that charcoal resulting from modern forest and heath fires bore a strong resemblance to fossil fusain. This observation led Harris to consider the role of fire in Tertiary times as an ecological determinant of vegetation, and as a selective factor in the evolution of Mesozoic flora. The dispersal of carbonaceous fragments through the air depends on settling velocity and hence on particle size. During a cruise across the Atlantic in 1980, Andreae³ filtered fine particles of carbonized material (2–5 µm in size)

from the atmosphere, and found that dispersal was so efficient that the fragment densities in remote parts of the Atlantic were comparable with those in rural continental areas. The presence of a 'soot' layer in clays from the Cretaceous/Tertiary boundary has subsequently raised the question of whether wildfire-generated smoke may have contributed to the change in climate that occurred during this period⁴.

Cope and Chaloner^{5,6} have used the presence of quantities of fusain in rocks as a basis for the reconstruction of palaeo-atmospheres, claiming that, since the Lower Carboniferous, atmospheric oxygen levels cannot have fallen below 30% of their present level (a figure subsequently revised to 60%). This conclusion is based on the oxygen levels that would have been required to support the extensive forest fires that generated the widespread influx of fusain into sediments.

Scott¹ has now extended the experimental work of Harris² by scanning electron microscopy of carbonized material generated from a variety of plant parts, such as wood, leaves and flowers (see figure). He finds very close agreement between the experimental and the fossil material in the preservation of detailed anatomical structure of the tissues. Charred material often retains a clear and almost perfect structure, especially in wood samples but also in more delicate organs. Reflectance microscopy has also proved a valuable tool. Wood charcoal and wood-



Micrographs of, left, a fusain fragment of a pteridosperm axis from the Lower Carboniferous (270 μm across) and, right, modern beech (*Fagus*) charcoal (125 μm across).

derived fusain have very high reflectance values, whereas charred leaf tissues have lower reflectance. There is also a positive correlation between the temperature of combustion and reflectance. It should, therefore, prove possible to interpret fossil fusains with some degree of precision, both in the location of their source and their mode of origin.

Charcoal in coal is derived from either local wetland forest fires or from a considerable input of charcoal from airborne sources. The possibility of water transport into the site of sedimentation is unlikely for many coals because the coal-forming mires seem to have arisen as ombrotrophic (rain-fed) systems, similar to modern raised bogs, rather than rheotrophic (stream-fed) wetlands. Scott suggests that the origin of fusain could be determined simply by taxonomic analysis. Gymnosperm material is likely to mean

upland fires and aerial input, whereas lycopod and fern fusain indicate local fire. The differential production and survival of carbonized particles from the different plants, must, however, be accounted for.

This line of research is promising both as a means of determining whether coal-forming mires were subject to fire (and, by implication, periodic drought), and also as a source of information on the ecological significance of fire in the evolution of plants and the development of vegetation.

Peter D. Moore is in the Division of Biosphere Sciences, King's College, Campden Hill Road, London W8 7AH, UK.

1. Scott, A.C. *Int. J. Coal Geol.* **12**, 443–475 (1989).
2. Harris, T.M. *J. Ecol.* **46**, 447–453 (1958).
3. Andreae, M.O. *Science* **220**, 1148–1151 (1983).
4. Wolbach, W.S., Lewis, R.S. & Anders, E. *Science* **230**, 167–170 (1985).
5. Cope, M.J. & Chaloner, W.G. *Nature* **283**, 647–649 (1980).
6. Cope, M.J. & Chaloner, W.G. *Nature* **290**, 428 (1981).

CHEMICAL SYNTHESIS

Scaling molecular Everests

John Mann

DURING a 2½ hour lecture in 1972, R. B. Woodward described the first total synthesis of vitamin B12. At that time it was the most complex naturally occurring compound that had succumbed to total synthesis. It had occupied the efforts of an army of chemists on two continents over a period of 12 years (R. B. Woodward *Pure appl. Chem.* **33**, 145–177; 1973; A. Eschenmoser & C. E. Wintner *Science* **196**, 1410–1420; 1977). In the audience that evening at Harvard was a young Japanese chemist, Yoshito Kishi, newly appointed to the staff of the Chemistry Department. He has now scaled his own 'chemical Everest' by synthesizing with his colleagues the structurally complex marine toxin, palytoxin (Y. Kishi *et al. J. Am. chem. Soc.* **111**, 7525–7530, 7530–7533; 1989).

Palytoxin, a product of marine coelenterates of the genus *Palythoa*, is the most toxic non-proteinaceous substance currently known. Its median lethal dose (LD50) is about 0.025 $\mu\text{g kg}^{-1}$ for the

rabbit and 0.45 $\mu\text{g kg}^{-1}$ for the mouse. Although the LD50 for man is not known, the toxin was used in the past by native Hawaiians as a spear poison.

Palytoxin was first isolated by the group of R. E. Moore and P. J. Scheuer in

1963. They estimated its relative molecular mass to be around 3,000 (it is actually 2678.5), but the molecular formula of $\text{C}_{129}\text{H}_{223}\text{N}_3\text{O}_{54}$ remained in doubt until the gross structure was finally established in 1981 (Fig. 1d). An initial problem was classification of the toxin. Most natural products contain repeating subunits derived from acetate or mevalonate or other small building blocks, but such was not the case with palytoxin. The best one can say is that it possesses a highly functionalized fatty acid chain interrupted by monosaccharide residues.

In the mid-1970s, the Hawaiian group (Moore and Scheuer) and a Japanese group led by Y. Hirata used chemical degradation and the newly emerging NMR and mass-spectroscopic techniques to determine the gross structure of palytoxin. As the complexity of the molecule emerged, it became clear that the toxin presented an unprecedented challenge for the synthetic chemist. Kishi accepted the challenge.

First, his group synthesized (1980–82) each of the fragments produced by degradation, thus providing an unambiguous confirmation of the gross structure, and also complementing the stereochemical assignments being made (on the basis of spectroscopic studies) by the other groups. By 1985, they had devised practical routes to eight key building blocks, all in stereochemically pure form; the stage was then set for development of methods for the convergent coupling of these fragments. The successful completion of these efforts is the subject of the two recent papers, and Kishi now has in hand palytoxin acid which lacks only the terminal N-(3'-hydroxypropyl)-3-aminopropenamide moiety of palytoxin.

There is no doubt that this synthesis could not have been accomplished in the 1970s. It required the revolution in the methods of asymmetric synthesis achieved in this decade, and the contemporaneous evolution of sophisticated two-dimensional NMR techniques to make it possible. The primary value of this latter

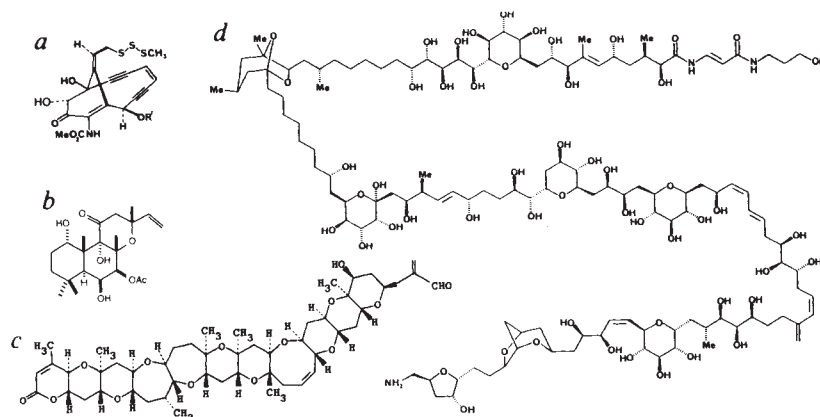


FIG. 1 Structures of palytoxin (d) and three other natural synthetic targets (see text).

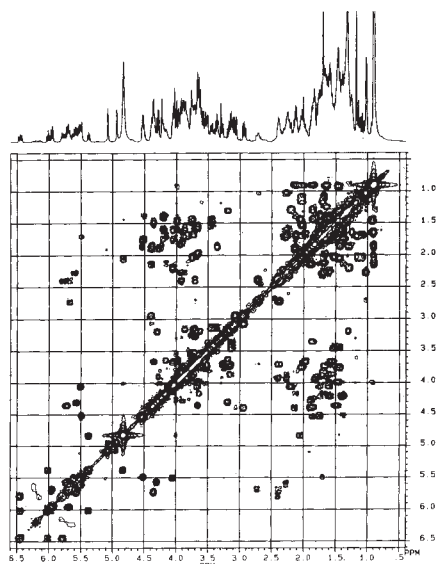


FIG. 2 Two-dimensional COSY NMR spectrum of palytoxin acid. Peaks along the diagonal are those found in a conventional NMR spectrum and indicate the local environment of each proton. Off-diagonal peaks emerge in the more elaborate method using two radio pulses and correlate nearby protons so that the molecular structure is better elucidated.

technique is that it allows the connections between carbon atoms or between molecular fragments to be delineated. The two-dimensional COSY (correlation spectroscopy) ^1H NMR spectrum (at 500 MHz) of palytoxin carboxylic acid reproduced in the second paper provides a cogent illustration of how the technique has revolutionized structure elucidation (Fig. 2).

As for the actual synthesis, two kinds of method are pre-eminent: the use of carbohydrates (and to a lesser extent monoterpenes) as stereochemically defined starting materials for synthesis; and the Sharpless epoxidation (a form of asymmetric induction) as a means of obtaining essentially stereochemically pure products from racemic starting materials. This desire to synthesize the natural stereoisomer of a molecule, rather than a racemic mixture of isomers, has been a consuming passion of most synthetic chemists during the past decade; and these are but two of the strategies for achieving this goal.

Kishi and co-workers also made much use of transition-metal catalysed coupling reactions between the various fragments, another area in which major advances have been made. It has to be said that the construction of these molecules was also greatly facilitated by older methods such as the Wittig reaction and the aldol reaction, albeit with greater control of stereospecificity than in the past.

So with the synthesis of palytoxin all but complete, the obvious questions are: was it worth the effort, and, what now? The answer to the first question has to be in the affirmative, because not only was a lot of new chemistry developed, but it was used

to construct a molecule with the correct stereochemistry at 64 asymmetric centres, and is thus tried and tested chemistry. As for the future, other daunting molecular targets are already under attack, for example the marine natural product brevetoxin B (Fig. 1c; K.C. Nicolaou, C.V.C. Prasad, P.K. Somers & C.-K. Hwang *J. Am. chem. Soc.* **111**, 6666–6690; 1989). But there are many much simpler structures that are popular targets for synthesis. Two particularly interesting ones are esperramycin A (Fig. 1a; see, for example, P. Magnus, R.T. Lewis and F. Bennett *J. chem. Soc. Chem. Commun.* 916–919; 1989) and forskolin (Fig. 1b; for a recent synthetic approach see G. Pattenden *et al. Tetrahedron* **45**, 5215–5246; 1989). Esperramycin A is a novel cytotoxic agent with a unique mode of

action; and forskolin, an activator of the enzyme adenylyl cyclase, is being evaluated for the treatment of hypertension, glaucoma, asthma and certain tumours.

On balance, the synthesis of relatively small, biologically active molecules is usually more rewarding. The target structures are generally more accessible, the cost is much more modest and, most importantly, analogues can be prepared for structure–activity studies within a reasonable time. The chances of obtaining a therapeutically useful molecule are thus enhanced, whilst there remains the opportunity to develop new, efficacious synthetic methods. □

John Mann is in the Department of Chemistry, University of Reading, PO Box 224, Reading RG6 2AD, UK.

LOWER-CRUSTAL PROCESSES

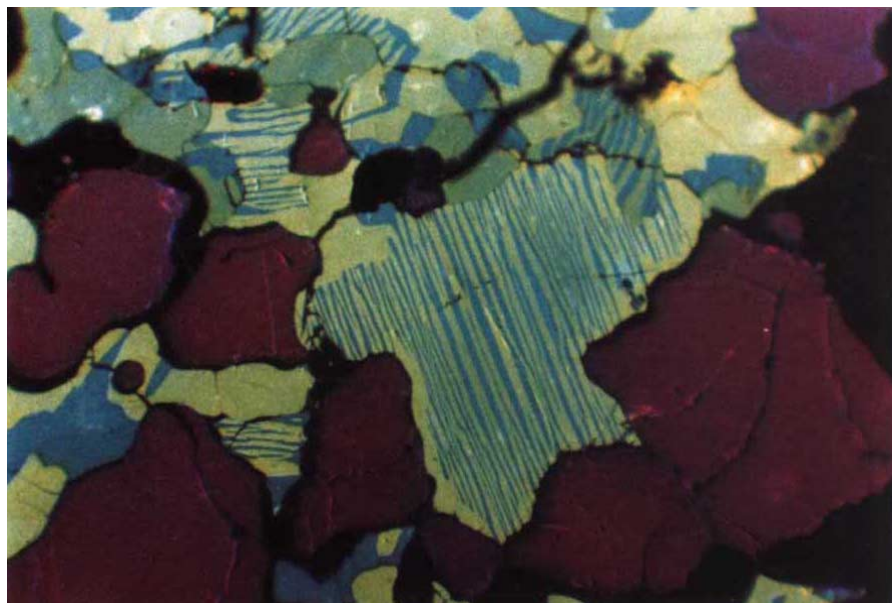
Granulites still being formed

J. D. Clemens

GEOLOGICAL opinion is deeply divided on the nature of the Earth's lower crust. Either, it is argued, the region is almost entirely composed of mafic (Mg–Fe-rich) meta-igneous rocks — or it is said that there is a wide variety of rock types. The rocks there are predominantly granulites — or granulites are only mid-crustal rocks. The lower crust is the main site of continental growth by basaltic underplating — or continents grow mainly by marginal accretion. Granulite formation is an equally contentious issue. Granulite-facies metamorphism is caused by deep burial of rocks in collision zones — or it is due to the introduction of mantle heat via mafic magma intrusion in extensional or

trans-tensional settings — or it is brought about by the influx of hot CO_2 -rich fluids. Granulites are ancient rocks — or they can be of essentially any age. Elsewhere in this issue, J.L. Hayob *et al.* (*Nature* **342**, 265–268; 1989) point the way to a resolution of some of these problems.

Hayob *et al.* have studied crustally derived metamorphic xenoliths retrieved from Quaternary volcanoes in central Mexico. The xenoliths originated as shaley sediments formed at the Earth's surface. But the authors' mineralogical evidence demonstrates that they have been subjected to temperatures of at least 950–1,110 °C (see figure) and pressures of 900–1,400 megapascals (corresponding



Cathodoluminescence photomicrograph of a granulite studied by Hayob *et al.*. The record high temperature at which the granulite was formed is revealed by the compositions of coexisting feldspar lamellae (blue and green stripes) formed during cooling. (Image 1.3 × 0.88 mm.)

Scott J. Carpenter

to lower-crustal depths in the region). Their data also imply that the metamorphism occurred relatively recently — no more than 30 million years ago. Hayob *et al.* interpret the distribution of xenolith localities to indicate that the granulite-grade metamorphic event was regional in extent (at least 100 km² in area and perhaps extending for more than a thousand kilometres to the north). The heat source is inferred to be underplated basaltic magma, small volumes of which reached the surface as the xenolith-bearing basanites. The authors emphasize their evidence for the lack of a protracted history of cooling of the xenoliths (to temperatures less than 900 °C), and the fact that basaltic magmatism continues to be active in the region. They use these observations to postulate that granulite-facies metamorphism continues even now in southern North America.

Hayob *et al.* suggest that the lower crust can contain supracrustal rock types; that, in this case, the lower crust is granulitic; that granulite-facies metamorphism can be powered by mantle heat through basaltic underplating away from mountain-building (orogenic) settings; and that these processes are still continuing. None of these ideas or general observations (except for the extreme temperatures and limited retrograde cooling) is uniquely that of the authors. Their contribution lies in documentation and in the recognition of the collective importance of the separate pieces of data. The authors' arguments

need some strengthening, particularly by the application of different age-dating techniques to xenolith mineral assemblages and integration with structural, tectonic and geophysical observations. But even as it stands, the evidence is sufficiently strong to make waves in the sea of geological opinion.

In the short term, the debate is unlikely to settle down. Granulite-facies metamorphism does occur over a broad range of conditions (700–1,100 °C; 400–1,500 megapascals; water activity 0–0.7) in varied tectonic settings. This suggests that a great variety of materials and processes are involved. The task ahead is to determine which are actually involved and which are most important. I expect that the underplating mechanism, combined with partial-fusion reactions in the absence of fluids, will turn out to be a principal factor in granulite-facies metamorphism, growth and differentiation of the crust, and the origin of upper-crustal granitic rocks. In an even broader context, I believe that core-mantle boundary instabilities, mantle plumes, anomalous conductivity in the lower crust, continental rifting, flood basalt outpourings and major biological events will all eventually come to be linked with these lower-crustal processes, operating on supracrustal rocks, in the same complex, discontinuous chain of events. □

J.D. Clemens is in the Department of Geology, Manchester University, Manchester M13 9PL, UK.

BIOMECHANICS

Gibbons swing stress away

R. McNeill Alexander

A FORCE acting on the end of a bone, parallel to its long axis, sets up stresses that are uniform across each cross-section. A force at right angles to the bone exerts bending moments on it, setting up tensile stresses in one face of the bone and compressive stress in the opposite face. When mammals run, the stresses in their bones are mainly due to bending moments. But when gibbons brachiate, swinging

loads. The radius, however, experiences bending moments because it is curved and is pulled on by the muscles that hold the fingers hooked over the branch.

A given force sets up much larger stresses in bones (or other long, slender structures) if it acts at even quite a small angle to the shaft than if it acts precisely axially. Gibbons can make do with remarkably slender arm bones because they load them axially: the humerus, for example, is 1.7 times as long as predicted for typical primates of the same body mass, but barely thicker than predicted².

Compare a gibbon of mass m and arms of length l , with a biped of the same mass and with legs also of length l . Suppose that the gibbon reaches out to a branch that is level with its shoulder and starts swinging. In the initial downswing it loses height l and gains speed v . The lost potential energy is converted to kinetic energy, so $v = \sqrt{2gl}$. The gibbon reaches this maximum speed in each swing, but its

mean speed over a complete cycle of brachiation depends on the angle of swing, α (see figure). For $\alpha = 90^\circ$, the mean speed is $0.5\sqrt{gl}$ (calculated as in ref. 3); for small values of α , it approaches $1.4\sqrt{gl}$. The force transmitted by the arm at the bottom of each swing is the weight plus the centripetal force, $mg + (mv^2/l) = 3mg$. If the arm is kept straight, the load on the bones is axial.

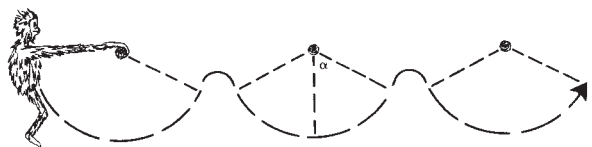
Our theoretical biped could maintain the axial load on its bones by keeping the supporting leg straight, as walking people do, but the axial load would be compressive rather than tensile as in the gibbon arm. Long leg bones could not be very slender or they would buckle under this load. Also, a biped using a straight-legged walk could not travel faster than $\sqrt{gl}$ because higher speeds would require downward accelerations greater than gravitational g (ref. 4). To travel faster, the biped would have to run, bending the legs at the knee and exposing the bones to bending moments. The peak force on a human runner's foot is about $2.8mg$, almost the same multiple of body weight as on the hand of our theoretical gibbon; but our theoretical biped would need leg bones thicker than the gibbon's arm bones to withstand the bending moments.

What would happen if the gibbon and the biped evolved longer limbs? By the argument already presented, the gibbon's speed would increase in proportion to the square root of its arm length. Similarly, the biped would increase speed in proportion to the square root of its leg length if it maintained dynamic similarity of gait⁵. In both cases the forces on the limb would remain unchanged, but the bending moments would increase in the biped in proportion to leg length. The longer-armed gibbon has no need for its bones to be thicker but the longer-legged biped does need thicker bones.

Gibbons are not limited by the speed they can gain in the first swing, but can accelerate by using leg movements like a child 'pumping' a swing⁶. They could accelerate faster by bending and extending the supporting arm instead of the legs, at appropriate stages of the swing, but that would result in bending moments. The penalty for relying on slender tensile arms is poor acceleration.

Finally, the really bad news, gibbon bones often get broken⁷. □

R. McNeill Alexander is in the Department of Pure and Applied Biology, University of Leeds, Leeds LS2 9JT, UK.



from branch to branch by their arms, the stresses in two of the principal arm bones are tensile right across the shaft, according to measurements reported by Sharon Swartz and her colleagues on page 270 of this issue¹. By attaching strain gauges to the bones of captive gibbons and recording, by radiotelemetry, the strains during brachiation, they found that the humerus and ulna experience almost axial tensile

1. Swartz, S.M., Bertram, J.E.A. & Biewender, A.A. *Nature* **342**, 270–272 (1989).
2. Aiello, L.C. *Symp. zool. Soc. Lond.* **48**, 331–358 (1981).
3. Preuschoft, H. & Demes, B. in *The Lesser Apes* (eds Preuschoft, H., Chivers, D.J., Brockelman, W.Y. & Creel, N.) 96–118 (Edinburgh University Press, 1984).
4. Alexander, R. *McN. Am. Sci.* **72**, 348–354 (1984).
5. Alexander, R. *McN. & Jayes, A.S. J. Zool. Lond.* **201**, 135–152 (1983).
6. Fleagle, J. *Nature* **248**, 259–260 (1974).
7. Schultz, A.H. *Bull. Hist. Med.* **7**, 571–582 (1939).

Parasitological teeth for evolutionary problems

Paul H. Harvey

EVOLUTIONARY biologists have much to gain from a closer association with parasitologists. It is not just that parasites have been overlooked for too long as a selective force, and that a considered assessment may show them to be at least partly responsible for a few of the problems that Darwin left us — sex, sexual selection and sociality have received attention recently. There is more to it than that. A parasitological perspective may shed light on evolutionary problems that do not necessarily involve one species parasitizing another. Graham Bell and Austin Burt's assessment of the evolutionary significance of B chromosomes, which was presented at a recent symposium on evolutionary approaches to parasitism*, nicely illustrates the point.

B chromosomes are distinguished from A chromosomes both by differing in number between individuals of the same species and by being unnecessary for the survival or reproduction of the individuals that possess them. They are extremely common, probably being present in 10–15 per cent of eukaryotic species, and are typically small. Often they are extensively heterochromatized — pointing to genetic inertness. Even though they may have originated from the autosomes or sex chromosomes of the lineage in which they are found, B chromosomes have clearly evolved properties of their own. In several plant species, B chromosomes are lost from the root tissue but retained in the aerial parts. And, in the growing shoots, they may be lost from tissues that will become nonreproductive, yet increase in numbers in tissue that will give rise to the germ line. Similarly, whereas small univalent chromosomes and chromosome fragments tend to be lost during meiosis, B chromosomes may actually accumulate. In organisms ranging from grasshoppers to rye grass, the B chromosomes move preferentially during meiosis to those daughter nuclei that will give rise to germ cells rather than polar bodies or non-generative nuclei.

Given this accumulation of B chromosomes in the germ cells, why do their numbers not increase from generation to generation? The answer seems to be that they reduce the darwinian fitness of their hosts. Cells with more DNA take longer to divide, and at least one study demonstrates that B chromatin adds about twice the amount of time to a typical cell division compared with a similar mass of A

chromatin. Morphological comparisons from a number of plant and animal species indicate that individuals with more B chromosomes usually exhibit delayed development, which is often accompanied by reduced vigour and fertility. In other words, B chromosomes bear all the hallmarks of successful parasites.

According to R. N. Jones and H. Rees (*B-Chromosomes*, Academic, London, 1982), the parasite theory for B chromosomes was suggested as long ago as 1945 by G. Östergren. Since then it has largely gone out of fashion and B chromosomes have been viewed as either of neutral selective value, or as being advantageous to the individuals or species that bear them. The parasite theory, however, does in fact seem successfully to accommodate recent observations. What is more, the theory makes fairly straightforward predictions that remain to be tested. For example, if B chromosomes are disadvantageous to their hosts, they should be lost from clonal lines and asexual species. In Bell and Burt's terminology, B chromosomes constitute not merely a disease, but a sexually transmitted disease.

The parasitological perspective on B chromosome evolution may lead to the reinterpretation of other observations. For example, there seems to be additive genetic variance for the presence and number of B chromosomes, which might be interpreted as differential resistance by particular sets of autosomes to invasion by B chromosomes. If that is so, the 'Red Queen' theory for the evolution of recombination (selection in response to ever-changing environmental challenges) might be invoked: that is, we might expect individuals possessing B chromosomes to increase recombination among the autosomes, so that new genetic variants that are resistant to B chromosomes would be more likely to arise. At the same time, we should expect successful B chromosomes to reduce the rates of recombination in their hosts.

So far, the evidence bearing on these questions, as reviewed by Bell and Burt, shows no consistent patterns. Working out concise predictions, suitable comparative tests and experimental protocols are tasks for the future. But to ask whether the work will be carried out by parasitologists or evolutionary biologists depends on an outmoded distinction between scientists with identical perspectives. □

Paul H. Harvey is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

Earth, wind and fire

SOME time ago the 'hydrogen economy' was proposed as a way of avoiding pollution. Hydrogen would be made in bulk by the nuclear-powered electrolysis of water or, perhaps, by the chemical transformation of oil. A universal fuel, it would burn cleanly to harmless water.

This green dream soon faded into well deserved oblivion. Nuclear power has its own potential for environmental disaster, and the generation of hydrogen from oil liberates as much greenhouse-inducing carbon dioxide as simply burning the oil in the usual way. But Daedalus is now reviving the idea. His plan is to turn the oil into hydrogen while it is still in the ground.

The modern steam-reforming catalytic hydrogen process would be hard to employ on crude oil *in situ*. Steam, heat and catalyst would all have to be fed continuously down the borehole. But Daedalus recalls the old partial-oxidation process which burnt crude oil in a limited supply of pressurized oxygen with a little steam in it, giving mainly hydrogen and carbon dioxide. No catalyst or added heat was needed. So Daedalus plans to pump oxygen and water down a test oil well, ignite the mixture electrically at the bottom and collect the resulting hydrogen from a central exit pipe.

Deep in the Earth, the process will run under great hydrostatic pressure. Carbon dioxide and undesirable by-products will dissolve in excess condensed water, and the pressurized solution will soak back into the porous oil-bearing rock as fast as the oil flows out of it. So the nasties will stay underground; only hydrogen will be volatile enough to escape to the surface.

Geohydrogen would be tricky to distribute. Overland, it could travel through the pipelines that now take oil and gas. But as bulk sea-freight it would be troublesome. Daedalus proposes to tow it round the world in captive balloons.

A small tug could easily tow a big balloon full of hydrogen. Even better, the balloon might tow the tug. The wind direction often varies markedly with height, and by winching the balloon up to the right altitude, the tug could be pulled in any chosen direction — a wonderful advance on the traditional sailing ship. Once discharged at its destination, the deflated balloon could be stowed on board the tug for the journey back to the oil field.

As a major fuel, hydrogen could be piped around like domestic or industrial gas. It could provide heat, power and clean water simultaneously and economically, wherever they are needed, ideally by means of a fuel cell. All potential pollutants would be left behind in the oil fields. Future geologists, prospecting for what remains of the oil, may be puzzled to strike dirty soda water.

David Jones

* British Society for Parasitology Autumn Symposium, City University, London, 29 September 1989. Organized by A. E. Keymer and A. F. Read.

Mechanism of solid-state fusion

SIR—Experiments by Derjaguin *et al.*^{1,2} indicating that MeV neutrons were emitted during the destruction of heavy ice or monocrystals of LiD went almost unnoticed, whereas recent claims of 'cold fusion'^{3,4} have attracted worldwide attention, ranging from highly enthusiastic confirmations to equally sceptical rejections. Here we wish to present some relevant considerations of mechanisms which could enhance nuclear interactions in the solid state.

A body of experimental evidence^{5,6} suggests that hydrogen resides in transition metals as a simple 'lattice gas'. Although the rate for nuclear fusion of $d + d$ in the lattice may be increased over the 'free' rate because of screening of the repulsive Coulomb potential by lattice electrons, the enhancement cannot be large enough⁷ to explain the claims of refs 3 and 4. Likewise, single-neutron tunnelling at low energies (the Oppenheimer-Philips mechanism⁸) will not do the job, and an excited resonance state in ^4He (ref. 9), because under these conditions the tunnelling time is much longer than the lifetime of the resonant state.

Any enhancement of the fusion rate must then be 'chemical' in origin — the potential barrier between deuterons may change shape substantially, or deuterons may gain some unusual thermal activation.

An increase in the density and effective mass of lattice electrons can cause a decrease of the shielding radius and a corresponding increase in barrier penetrability. In addition, the combined effects of a redistribution of electrons and lattice deformation may produce a local minimum in the potential barrier. Under the right conditions, a quasi-stationary complex can form, and an effect known as resonance transparency may appear¹⁰. If such a complex has a vibrational frequency with an energy ~ 0.1 eV, the effective collision frequency of deuterons can be greater by as much as a factor of 10^9 over the simple 'lattice gas' value.

A considerable increase in the fusion rate would also come about if a single deuteron were able to acquire ~ 10 keV of energy. In a lattice in equilibrium such

an energy transfer would not occur, but perhaps during the destruction of a lattice there might be significant localization of the destruction energy, which could be transferred to absorbed deuterons. A macroscopic version of this effect is found in the electrostatic acceleration of deuterium ions in cracks, which, as noted by Cohen and Davies in their recent News and Views article¹¹, is apparently responsible for the appearance of neutrons during the shock destruction of D_2O and LiD (refs 1 and 2).

This is then a problem in piconuclear (pressure-induced) reactions rather than in thermonuclear reactions. Common features may apply in the various experiments: high overvoltages and D compression in ref. 3, cold fusion in natural geophysical conditions in ref. 4, and high static pressures and shears in refs 1 and 2.

VITALII I. GOLDANSKII
FYODOR I. DALIDCHIK

*N.N. Semenov Institute of Chemical Physics,
USSR Academy of Sciences,
Ulitsa Kosygina 4,
117334 Moscow,
USSR*

Sex in diploids

SIR—Kirkpatrick and Jenkins¹ suggest that the origin and maintenance of sex in diploids is related to the advantage conferred by genetic segregation. They contend that the segregation advantage gained by facilitating the fixation of advantageous mutations more than compensates for the twofold reproduction advantage of asexual reproduction. We point out here that the segregation advantage they discuss is present even when only a very small proportion of individuals in a population reproduce sexually.

To illustrate this, we have compared the pattern of increase in the frequency of a

small amount of sexual reproduction, the continued increase of the advantageous allele from 0.5 to fixation is nearly as fast as that in a completely sexually reproducing population.

Kirkpatrick and Jenkins show that there is a plateau at a frequency of 0.5 for the heterozygous asexual population while it is waiting for the occurrence of a second mutation. We observe a temporary plateau at an allelic frequency of 0.5 only when the proportion of sexual reproduction is about 10^{-6} or smaller. This plateau, which is generally very short and in the most extreme case in the figure only doubles the time to fixation of the favourable mutant, occurs because the population consists of nearly 100% heterozygotes at

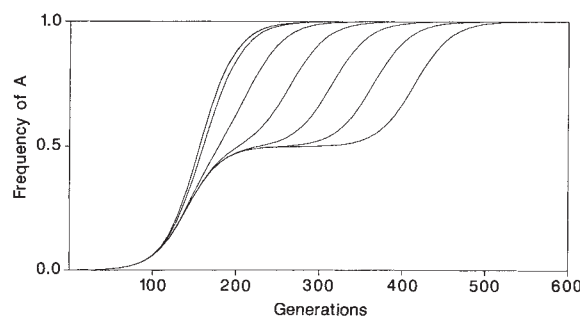
the beginning of this period and is waiting for segregation to produce AA homozygotes (which it will at a rate about $1/4$ times the rate of sexual reproduction). We also find that the relative duration of this plateau is even shorter when the selective difference between the homozygotes is reduced.

A number of species have a small (but not zero) amount of sexual reproduction^{2,3}. Perhaps this is a way in which a species can have both the segregation advantage discussed by Kirkpatrick and Jenkins and the twofold advantage of asexuality many others have noted. In other

words, we see why some sexual reproduction should be favoured over complete asexual reproduction because of the segregation advantage, but it is not clear why a lot of sexual reproduction should be favoured over a little⁴.

PHILIP W. HEDRICK
THOMAS S. WHITTAM

*Department of Biology,
Pennsylvania State University,
University Park,
Pennsylvania 16802,
USA*



The change in frequency of advantageous mutation A for different amounts of sexual reproduction. For the curve on the far right the proportion of sexual reproduction is 10^{-7} (1 in 10 million individuals reproduce sexually). Successive curves to the left are for increasing amounts of sexual reproduction (10^{-6} , 10^{-5} , ..., 10^{-1}). The curve for 10 per cent sexual reproduction is indistinguishable from 100 per cent sexual reproduction.

favourable allele in a population with only sexual reproduction to that in populations where there is a mixture of sexual and asexual reproduction². As an example, the figure shows the change in allelic frequency of advantageous mutation A when there is a 10 per cent selective difference between homozygotes and the heterozygotes are exactly intermediate in fitness. Note first that the initial increase from 0.001 to nearly 0.5 occurs essentially at the same rate for all levels of sexual reproduction. Furthermore, even if there is only a very

- Derjaguin, B.V. *et al. Sov. Colloid J.* **48**, 12 (1986).
- Kluyev, V.A. *et al. Sov. J. tech. Phys. Lett.* **12**, 1333 (1986).
- Fleischmann, M., Pons, B.S. & Hawkins, M. *J. electroanal. Chem.* **261**, 301 (1989); and erratum, **263**, 187–188 (1989).
- Jones, S.E. *et al. Nature* **338**, 737 (1989).
- Gel'd, P.V. *et al. Hydrogen and Physical Properties of Metals and Alloys* (Nauka, Moscow, 1985).
- Hydrogen in Metals* (eds Alefeld, G. & Völkl, I.) 2 vols (Springer, Berlin, 1978).
- Baym, G. & Leggett, A.J. *Nature* **340**, 45 (1989).
- Oppenheimer, J.R. & Phillips, M. *Phys. Rev.* **48**, 500 (1935).
- Bevalacqua, J. *Phys. Rev. C* **27**, 2417 (1983).
- Bohm, D. *Quantum Theory* (Prentice-Hall, New York, 1952).
- Cohen, J.S. & Davies, J.D. *Nature* **338**, 705–706 (1989).

SIR—I believe that the mechanism proposed by Kirkpatrick and Jenkins¹ is unlikely to offer a sufficient explanation for the maintenance of diploid sex.

Of critical importance in assessing the potential role of the segregation advantage¹ in maintaining diploid sex is the fact that the gradual accumulation of beneficial homozygous loci by a sexual population can only lead to gradual increases in its relative fitness. Hence, a significant time period (Kirkpatrick and Jenkins provide an estimate of $\sim 10^6$ generations) would be required before the level of homozygosity in a sexual population could counterbalance the supposedly inherent twofold fitness advantage of asexual reproduction. But throughout these many generations the asexual population would enjoy a large, slowly diminishing selective advantage, which should result in the competitive elimination of the sexual population long before its fitness approaches that of the asexual population. In addition, competition with a more fit asexual population would lead to a continuous reduction in the size of the sexual population, slowing its relative rate of production of new beneficial mutant alleles, and further reducing its competitive prospects. Even if the sexual population eliminated the asexual population, new asexual mutants, starting with the same genetic complement as the sexual population, would again enjoy a selective advantage for, say, $\sim 10^6$ generations. In the absence of any advantage to sex other than an initially higher rate of fixing homozygous loci, the repeated appearance of such mutants would keep a sexual population at a constant disadvantage.

The segregation advantage cannot explain the evolution of sex or its maintenance in haploid forms¹. However, the existence of sex in species with predominantly haploid life cycles implies a selective advantage sufficient to overcome any costs associated with meiosis. Surely, parsimony supports a single explanation for the maintenance of sex in both diploid and haploid species rather than hypotheses exclusive to either group.

ROBERT L. DUNBRACK

Department of Biology,
Memorial University of Newfoundland,
St. John's, Newfoundland,
A1B 3X9 Canada

KIRKPATRICK AND JENKINS REPLY—We showed¹ that the process of genetic segregation can generate an evolutionary advantage to sexual reproduction that may help explain its prevalence over asexual reproduction. Hedrick and Whittam point out that the segregation advantage we identified will be almost as effective in a population in which most individuals are asexual and only a small fraction are sexual. Such a facultatively sexual population would have the best of both worlds:

most of the segregation advantage from sexual reproduction and most of the twofold reproductive advantage from asexual reproduction. Species of this type, however, are virtually unknown among the animals⁵, and so this may not be an evolutionary option. (We do not consider automixis (selfing) here because the ubiquitous deleterious effects of inbreeding⁶ probably select against it in most populations.) We have extended Hedrick and Whittam's observation by considering the more common (though still rare) situation of cyclic parthenogenesis, in which the population undergoes a number of strictly asexual generations followed by a single generation of sexual reproduction⁵. The results are similar to Hedrick and Whittam's: an occasional generation of sexual reproduction is sufficient to give most of the segregation advantage and still retain most of the twofold reproductive advantage of asexuality.

Why then are facultative and cyclic parthenogenesis uncommon? We have two possible explanations. First, the segregation effect we identified is only one of the evolutionary advantages to sex. Others, such as effects from recombination^{4,5,7}, may work in conjunction with the segregation effect to give purely sexual reproduction an overall advantage. Second, constraints may preclude facultative and cyclic parthenogenesis as evolutionary options for many taxa^{4,5}. Pure apomictic parthenogenesis has appeared dozens of times among the animals, but

facultative and cyclic apomictic reproduction have arisen perhaps only half a dozen times⁵. Thus, for most taxa the evolutionary possibilities may be purely sexual or purely asexual reproduction. In this case, the benefits to sex from the effects of segregation and recombination may be sufficient to maintain sexual reproduction.

Dunbrack reiterates two points made in our paper: that there can be a substantial time lag before the segregation advantage offsets the twofold reproductive advantage of asexual reproduction; and that the segregation advantage will be greatly weakened when much or all of the genome is selected in a haploid state for a substantial part of the life cycle. Dunbrack feels these points are fatal to our hypothesis. We disagree based on the arguments presented in our paper, and re-emphasize that there need not be only one evolutionary advantage to sex — the segregation effect can be augmented by others.

MARK KIRKPATRICK
CHERYL D. JENKINS

Department of Zoology,
University of Texas,
Austin, Texas 78701, USA

1. Kirkpatrick, M. & Jenkins, C.D. *Nature* **339**, 300 (1989).
2. Marshall, D.R. & Weir, B.S. *Heredity* **42**, 159–172 (1979).
3. Jackson, J.B.C. et al. *Population Biology and Evolution of Clonal Organisms* (Yale University Press, 1985).
4. Maynard Smith, J. *The Evolution of Sex* (Cambridge University Press, 1978).
5. Bell, G. *The Masterpiece of Nature: The Evolution and Genetics of Sexuality* (University of California Press, Berkeley, 1982).
6. Charlesworth, D. & Charlesworth, B. A. *Rev. Ecol. Syst.* **18**, 237–268 (1987).
7. Michod, R.E. & Levin, B.R. *The Evolution of Sex: An Examination of Current Ideas* (Sinauer, Sunderland, 1988).

More on cold fusion

SIR—Processes involving cosmic-ray muons have been proposed^{1–3} to explain the small rate of fusion neutrons observed⁴ in the cold-fusion experiments. Although fusion neutrons have not been observed in electrochemical cells placed in muon beams^{5,6}, the flux from the control experiment has not been measured. Any cold-fusion neutrons of amuonic origin would remain unaffected by the muon beam. These muon experiments therefore demonstrate negative results for cold fusion — both muonic and amuonic in deuterium-charged metals. The muon experiments establish that the deuteron concentration in the samples does not reach abnormally high densities because the muons are lost to the more electro-positive host-metal ions as expected for normal ratios of deuterons to metal-ions (up to two). Under conditions of high-density non-equilibrium deuteron packing as predicted by some of the cold-fusion experiments⁷ and theories, the muon kinetics are not known, but loss to the host metal should be hindered, as it is known to be sometimes delayed⁸ even in ordinary compounds because of intermediate molecular states.

Known cosmic-ray muon fluxes^{9,10}, their

stopping rates¹¹, and established muon-catalysed-fusion parameters^{12,13} indicate that exposure of a large enough volume of liquid deuterium (about 1 litre) to cosmic-ray muons should also yield a low-level neutron count of about 10^{-2} s^{-1} . There is a possibility of using these fusions for fissile-fuel breeding. Although the rate would be extremely slow, the muons would be free and the muon-production cost factor could be dropped from the efficiency calculations for muon-catalysed fusion.

LALI CHATTERJEE

Physics Department,
Jadavpur University,
Calcutta 32,
India

1. McEvoy, A.J. & O'Sullivan, C.T.D. *Nature* **338**, 711 (1989).
2. Chatterjee, L. Preprint JU HEP/89/2, Jadavpur Univ. (1989).
3. Chatterjee, L. *Proc. ICENES 1989* (in the press).
4. Jones, S.E. et al. *Nature* **228**, 737–740 (1989).
5. Nagamine, K. *Proc. Conf. Muon-Catalysed Fusion* (Oxford, 1989) (in the press).
6. Davies, J. *Proc. Conf. Muon-Catalysed Fusion* (Oxford, 1989) (in the press).
7. Fleischmann, M. & Pons, S. *J. electroanal. Chem.* **261**, 301–308 (1989); and erratum, **263**, 187–188 (1989).
8. Gerstein, S.S. & Ponomarev, L. in *Muon Physics* Vol. 3 (eds Hughes, V.W. & Wu, C.S.) (Academic, New York, 1978).
9. Hayakawa, S. in *Cosmic Ray Physics* Vol. 22 (ed. Marshak R.E.) (Wiley Interscience, New York, 1969).
10. Yashpal & Peters, B. *Nat. fys. med. Dan. Vit. S.* **33** (1964).
11. Wightman, A.S. *Phys. Rev.* **77**, 521 (1950).
12. Jones, S.E. *Nature* **321**, 127–133 (1986).
13. Chatterjee, L. *Phys. Lett.* **A137**, 4–6 (1989).

High-latitude ozone loss outside the Antarctic ozone hole

M. H. Proffitt^{*†}, D. W. Fahey^{*}, K. K. Kelly^{*} & A. F. Tuck^{*}

^{*}NOAA Aeronomy Laboratory, 325 Broadway, Boulder, Colorado 80303, USA

[†]CIRES, University of Colorado, Boulder, Colorado 80309, USA

Data taken during the 1987 Antarctic Airborne Ozone Experiment based in Punta Arenas, Chile, are used to show that from mid-August until the end of the mission in late September there was a high-latitude ozone loss outside the Antarctic ozone hole. Therefore, not only is the geographic extent of the ozone loss larger than that generally identified as chemically perturbed, but ozone is lost earlier in the year than previously reported. These results, when compared with long-term temporal trends of column ozone, indicate a possible anthropogenic component for this loss.

MEASUREMENTS from the past few years have shown that stratospheric ozone decreases substantially over Antarctica during September and October and that these decreases have intensified dramatically during the past decade. The intensification has been attributed to man's release of certain chlorine-containing compounds, primarily chlorofluorocarbons. The region of very high ozone decrease, usually referred to as the ozone hole, approximately covers the Antarctic continent, and is coincident with what is commonly called the polar vortex. Horizontally, the polar vortex is roughly defined by its perimeter, the polar night jet, a circumpolar wind belt in the stratosphere that seems to restrict exchange of air through the perimeter. The seasonal period of low ozone values is observed to terminate with the annual dissipation of the Antarctic vortex during the spring warming, at which time the air within the vortex is no longer restricted but allowed to mix freely with ozone-rich air from other latitudes. Transport towards the Equator and mixing of ozone-poor air with air at lower latitudes result in localized decreases of ozone¹. The extent and intensity of these decreases are of global concern.

The seasonal decrease in ozone is large enough to provide an unequivocal signature of loss that is easily distinguished from natural short-term dynamical variability, and has been verified from the ground, aircraft, balloons and satellites. In this paper 'ozone loss' is used to identify a change in ozone due to chemical mechanisms (rather than dynamical processes) that is not predicted within standard gas-phase photochemistry models. We do not intend to imply the loss is necessarily due solely to chlorine. Measurements of total column ozone (total number of ozone molecules vertically above some area of the Earth) made from the ground and from a single station first identified ozone loss over Antarctica². Satellite column measurements then verified that it was not just a local phenomenon³. Measurements obtained from balloons provided snapshots of the vertical distribution of ozone at a few Antarctic locations and presented striking accounts of large ozone decreases when comparing profiles from August with those from October^{4,5}. But the changes observed were not simply a gradual deterioration of the ozone layer, but included a substantial day-to-day variability and frequent horizontal layering. This means that the region of decreased ozone is neither vertically nor horizontally homogeneous, and that quantification of the altitude-dependent

decrease of ozone requires good temporal and geographic coverage. Although satellite measurements of total column ozone are frequently available, they cannot identify the altitude dependence of the ozone loss. *In situ* aircraft measurements can provide vertical profiles and extended horizontal coverage, but they are still insufficient to sample the entire geographic extent of the vortex frequently.

Here we present a different approach to estimating the ozone (O_3) loss inside and, more interestingly, outside the ozone hole—one that completely avoids the problems of sampling an inhomogeneous air mass. As standard gas-phase photochemistry is very slow in the region of this analysis (that is, during high-latitude winter in the lower stratosphere where O_3 has a photochemical replacement time of more than a year^{6,7}), O_3 can be used effectively as a conservative tracer. The sum of the reactive nitrogen species (abbreviated as NO_y , and including in that sum NO , NO_2 , NO_3 , HNO_3 , N_2O_5 and $ClONO_2$) and nitrous oxide (N_2O) are also conserved in this environment. By combining *in situ* measurements of these tracers with empirically derived formulations that are supported by two-dimensional atmospheric models, minimum reference values for unperturbed O_3 can be estimated for any air parcel sampled. This reference value for O_3 can then be compared with the measured value, and the difference is a deviation from conservative behaviour that can be interpreted as a chemically induced change not within the standard gas-phase photochemistry. The formulations presented are first applied outside the region usually referred to as chemically perturbed. Then the analysis is used in a different manner, but this time both outside and inside. It is assumed for the analysis that there has been sufficient simultaneous sampling of the trace constituents NO_y , N_2O and O_3 ,

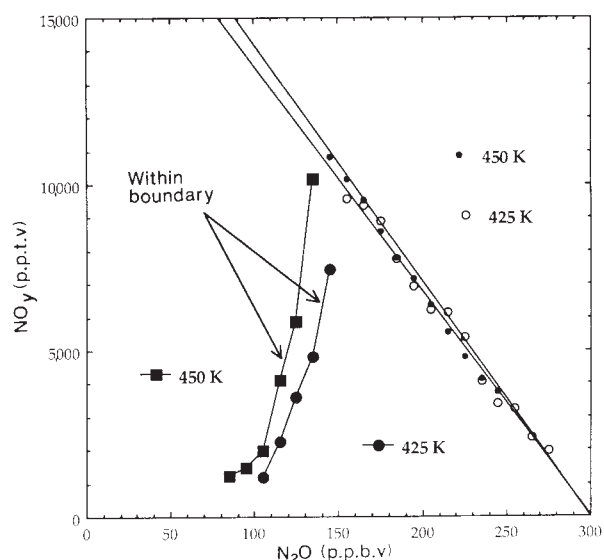


FIG. 1 NO_y against N_2O from 1987 AAOE for the 425-K and 450-K flight levels and averaged for the seven flights where data are available over intervals of 1° latitude. Data taken within clouds were excluded to obtain only gas-phase NO_y . Linear least-squares fits to the data outside the boundary are shown for each flight level.

spatially and temporally, to determine reliable reference formulations.

To discuss accurately the morphology of the ozone hole, the concept of the boundary of the chemically perturbed region is used as introduced and developed by Proffitt *et al.*⁸ from data taken during the 1987 Antarctic Airborne Ozone Experiment (AAOE). The boundary, defined by a level of chlorine monoxide ($\text{ClO} = 130$ p.p.t.v. (parts per 10^{12} by volume)), effectively locates where this critical component in the catalytic destruction of O_3 by chlorine has a sharp poleward increase. It was found that there were also dramatic poleward decreases in NO_y and H_2O at this boundary, hence the basis for describing the interior region as chemically perturbed. We shall refine this description slightly by referring to this region as 'high chemically perturbed', and thereby not exclude a possible lesser degree of chemical perturbation outside the boundary. Proffitt *et al.* have shown that this boundary is slightly poleward of the peak wind speeds of the polar night jet and that it coincides with the latitude of the strong poleward decrease in total column O_3 as measured by satellite. In addition, the boundary locates the perimeter of the large isentropic temporal O_3 decrease (ozone hole) to within at least 1° of latitude (~ 111 km).

Danielsen and Houben⁹ give heuristic arguments for winter-time diabatic cooling over Antarctica persisting into early spring, carrying ozone-rich mid-latitude air downward and poleward into the polar vortex. Proffitt *et al.*¹⁰ have presented supporting evidence from AAOE indicating a slow progression of diabatically cooling and descending lower stratospheric air advectively spiralling poleward as it crosses the polar jet into the ozone hole. This view can be compared with that of Tuck¹¹, who also concludes that there is a significant mass flow through the vortex, but argues that it is the downward diabatic motion within and outside the vortex that is important. With either view, the region near and within the ozone hole is being slowly replenished with mid-latitude ozone-rich stratospheric air during the period of O_3 decrease through diabatic descent. If this is the case, temporal trends in measured O_3 will not indicate the total loss of O_3 in a particular parcel of air unless the time series begins before there is any loss and the measurements follow the descending parcel throughout the time interval of the trend¹⁰. This is a difficult, if not impossible, sampling requirement, even with precise knowledge of a parcel's descent path. Possible mixing of air parcels of dissimilar origin during or following the O_3 destruction process further complicates the sampling problem.

The analysis presented is based entirely on the *in situ* data taken during AAOE from the high-altitude ER-2 aircraft. During the mission the average latitude of the boundary was 66°S , but varied from 59°S to 71°S , its position responding to synoptic-scale meteorological forcing of the polar vortex. The data are generally restricted vertically to the two levels of 425 ± 10 K and 450 ± 10 K in potential temperature (corresponding to geometric altitudes ~ 17.5 km and 19 km respectively over Antarctica) and horizontally from 53°S to 72°S latitude at $\sim 65^\circ\text{W}$ longitude. The longitudinal coverage is quite restricted, extending only from the southern tip of South America to the region over or near the Antarctic Peninsula. Also, data were collected during the transit flights between Moffett Field, California (37°N) and Punta Arenas, Chile (53°S), where the twelve ER-2 Antarctic flights originated.

Lower limit for unperturbed ozone

The stratospheric source for NO_y is the very unreactive species N_2O . The primary destruction mechanism for N_2O is ultraviolet radiation at wavelengths ≤ 225 nm, has N_2 and $\text{O}(^1\text{D})$ as principal products, and occurs primarily in the middle of the stratosphere above 30 km. A secondary loss occurs also in the middle stratosphere where NO is created in the reaction of $\text{O}(^1\text{D})$ with N_2O , providing the primary source of NO_y . Plots of N_2O against NO_y outside the boundary show them to be linearly related (Fig. 1). These data are from the two flight levels of 425 K and

450 K of potential temperature and were analysed separately by averaging within 1° -latitude bins with respect to the boundary and over a one-month period. The data in Fig. 1 labelled as within the boundary (corresponding to $\text{N}_2\text{O} \leq 140$ p.p.b.v. (parts per 10^9 by volume)) are known to be from within the highly chemically perturbed region. Linear least-squares fits are shown for the data outside this region and it is apparent that the straight line that fits the data at the lower altitude (or potential temperature) also fits the data at the upper altitude rather well. The linear fit to all of the points in Fig. 1 outside the boundary is given by

$$[\text{NO}_y] = 20,800 - (69.8 \times [\text{N}_2\text{O}]) \quad (1)$$

where the mixing ratio for NO_y is in p.p.t.v and for N_2O is in p.p.b.v.^{10,12,13}. It is noteworthy that data taken during the flights from Punta Arenas to Moffett Field, where minimum values for N_2O of 210 p.p.b.v. were observed, also fall on the same straight line¹⁰. There are few further data available to verify the persistence of the linear relationship at other seasons, latitudes and altitudes, although the mid-latitude Southern Hemisphere balloon data taken by Galbally *et al.*¹⁴ up to 30 km have the same characteristic. This persistent linear relationship and lack of curvature results from a balance of photochemistry, transport

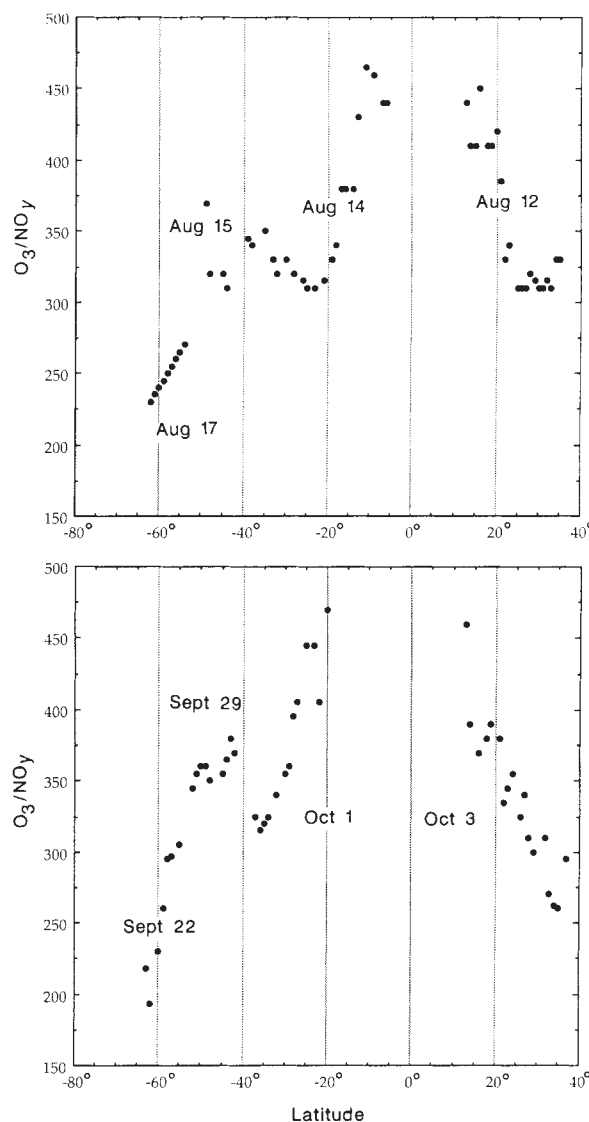


FIG. 2 O_3/NO_y latitude survey from a, 12 August to 17 August, 1987 and b, 22 September to 3 October, 1987.

and mixing. The N_2O intercept and slope are probably determined in the tropics where upper tropospheric and lower stratospheric air with N_2O of ~ 300 p.p.b.v. and $NO_y < 0.5$ p.p.b.v. (intercept) rises into the tropical middle stratosphere where NO_y is produced from N_2O apparently with an average efficiency of $\sim 7\%$ (slope). Mixing, as the middle stratospheric air proceeds poleward, probably contributes to the observed small variability from linearity observed at high latitudes.

The ratio of O_3 to NO_y (O_3/NO_y) can now be used to bring O_3 into the analysis. Figure 2a is a plot of O_3/NO_y against latitude taken at the start of the mission and includes data taken in transit from Moffett Field to Punta Arenas. The ratio is > 300 over a wide range in latitudes, but decreases at the most southerly latitudes. Figure 2b is a similar latitude survey including the last Antarctic flight and the transit flight data during the return to Moffett Field at the end of the mission. Still lower values at the southernmost latitudes and a few data points below 300 at about $35^\circ N$ are the primary differences between the two surveys. Both figures are restricted to pressure altitudes from between 18 km and 21 km, and all data within the boundary are excluded. In both plots $O_3/NO_y > 300$ at all latitudes north of $\sim 55^\circ S$, except on the flight of 3 October over California, where a few values as low as 260 are found. Both surveys also show an enhanced ratio in the tropics where, at 20 km, stratospheric air is apparently encountered that had recent O_3 production without the equivalent NO_y production required for the usual ratio found at mid-latitudes. Other latitude surveys of ER-2 data taken during the Stratosphere Troposphere Exchange Project in January and February 1987 from $40^\circ N$ to $30^\circ S$ have been analysed (D. M. Murphy *et al.*, manuscript in preparation) and show a minimum for the ratio of 280 when restricting the data to potential temperatures > 430 K. Also presented by Murphy *et al.* are two-dimensional model calculations^{15,16} and satellite data¹⁷ in the tropics and mid-latitudes covering pressures down to 7.5 mbar. Both the models and the data show that the O_3/NO_y ratio is > 300 in the lower stratosphere and indicate either no vertical change or a slight decrease with altitude. Qualitatively these observations are consistent with the well-known mid-latitude mixing ratio maximum for O_3 and NO_y at 30–35 km with the maximum for NO_y a few kilometres higher than the O_3 maximum. These maxima correspond to N_2O values of 30–100 p.p.b.v., values much less than those measured outside the boundary and, as seen in Fig. 1, generally less than any data from AAOE.

From the observations made during ER-2 latitude surveys it is reasonable to assert that if NO_y and O_3 are given in identical units (such as p.p.b.v.) the inequality

$$[O_3]/[NO_y] \geq 270 \quad (2)$$

is representative of late winter in the Southern Hemisphere north of the boundary. The data and model calculations indicate that a larger value of 300 for the minimum ratio could be justified. But there is no established record of simultaneous measurements of NO_y and O_3 covering all latitudes, seasons and altitudes of interest from which unequivocal formulations relating O_3 to NO_y could be derived. So the conservative approach taken here is appropriate, although it may underestimate the O_3 losses that have occurred.

Ozone losses

On 17 August between $54^\circ S$ and $63^\circ S$ (Fig. 2a), there was a significant difference between the expected O_3/NO_y ratio (inequality (2)) and its measured value. As NO_y was not perturbed and O_3/NO_y is less than our reference level of 270, an O_3 loss is indicated in that region. The value for O_3/NO_y of 230 at $63^\circ S$ yields a 15% loss of O_3 outside the highly chemically perturbed region on 17 August. Data from the flight of 23 August also indicates a 15% loss outside the boundary, and the flights of 28 August and 22 September show a larger loss of 30%. But rather than giving flight-by-flight analysis, Fig. 3 shows averages

for the period 23 August to 22 September to summarize the results and demonstrate that the O_3 losses are not isolated events found on certain flights, but were prevalent throughout the period and clearly seen in the averages. This is achieved by referencing the flight-leg data for NO_y and O_3 relative to the boundary, and then averaging the O_3/NO_y ratio over bins of 1° latitude. Only data for which NO_y is believed not to be enhanced by particles (as indicated by simultaneous particle measurements) have been plotted. This averaging procedure is done independently for both of the flight levels. The vertical bars represent the sample standard deviation and indicate how flight-by-flight variability affects the results.

The abrupt increase south of the boundary is due to the high degree of denitrification within the highly chemically perturbed region¹³ and provides little information on the O_3 loss. We deal below with this region in a different way. However, outside the boundary where the NO_y is known to be unperturbed relative to N_2O , there are depressed ratios even below 200 near the boundary. These low ratios correspond to an average minimum O_3 loss of 15–30% in the first 5° of latitude outside the highly chemically perturbed region at these levels. It should be noted that the averaging over the one-month period is in a region where diabatic descent is believed to be taking place. An important feature in Fig. 3 is that outside the boundary the ratios at the 450-K level are $\sim 20\%$ higher than those at the 425-K level. This is not what would be expected from the model calculations and satellite data mentioned above, which show either no change in O_3/NO_y with altitude or a decrease. The implication is that the lowest values seen near but outside the boundary could not have come simply from vertical descent in this region. Proffitt *et al.*¹⁰ show that within the averaging procedure used for Fig. 1, $< 10\%$ of the air outside the boundary had come recently from the highly chemically perturbed region. Thus, if these results are accepted (that is, ongoing diabatic descent and little outward mixing at these potential temperatures) along with inequality (2), the decrease in the O_3/NO_y ratio, seen as the boundary is approached from the north, cannot be due to transport or mixing of air vertically nor horizontally, and therefore must be caused by some chemical mechanism operating at these levels outside the highly chemically perturbed region.

The isentropic temporal changes in O_3 outside the boundary

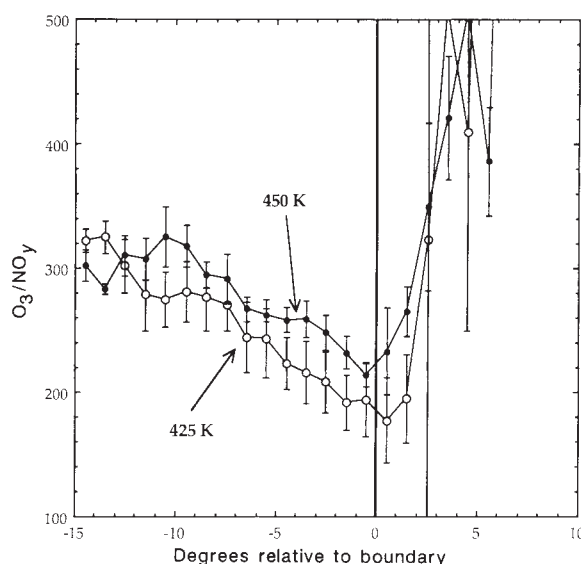


FIG. 3 O_3/NO_y against latitude relative to the boundary (heavy vertical line) for the 425-K and 450-K flight levels and averaged for the seven flights where data are available over intervals of 1° latitude. Data taken within clouds were excluded to obtain only gas-phase NO_y . Sample standard deviations are given as vertical bars.

during this one-month period are small in comparison with those inside, and in general are not significantly different from zero¹⁸. As discussed in the introduction, losses of O₃ in a parcel of air undergoing diabatic descent will not be represented accurately by a temporal trend calculated at a constant potential temperature, because the losses will be masked by the usual vertical gradient in O₃. Therefore the lack of a significant isentropic temporal trend outside the boundary in no way indicates that no loss of O₃ is occurring there.

To extend our analysis into the ozone hole and provide a check for consistency between two methods, equation (1) and inequality (2) are now combined into an inequality relating unperturbed O₃ directly to N₂O. If O₃ and N₂O are both in p.p.b.v.

$$[\text{O}_3] \geq 5616 - (18.85 \times [\text{N}_2\text{O}]) \quad (3)$$

provides a minimum value for unperturbed O₃. First note that this formula is based on measured data and may reflect systematic errors in the measurements. In this regard, the O₃ measurement used here was generally within 2% of an independent measurement of O₃ made simultaneously on the ER-2^{19,20}. There was no independent check of the NO_y measurements except as they related to the O₃ data over a long history of more than 50 ER-2 flights, but they are reported to be accurate to within $\pm 20\%$ and repeatable (precise) to within at least 0.1 p.p.b.v.¹³. The values used for N₂O are from the ATLAS instrument^{21,22} and include a recent recalibration of the data that was not included in earlier papers^{8,10-13,21,22}. The reported accuracy of the N₂O measurements is $\pm 10\%$.

There is more uncertainty in applying inequality (3) inside the boundary where N₂O < 140 p.p.b.v. (see Fig. 1), stemming from our lack of information to support equation (1) and inequality (2) at these more southerly latitudes. Outside the boundary, equation (1) was used to test where NO_y was perturbed, then inequality (2) was used to predict O₃. By extending the application of inequality (3) to 72° S with its lower values for N₂O it is assumed that equation (1) remains linear and inequality (2) is still representative of O₃/NO_y. The use of equation (1) at lower values of N₂O is supported by noting that the linear relationship between NO_y and N₂O is predicted to persist at mid-latitudes to at least 14 p.p.b.v. of NO_y in both a one-dimensional radiative convective photochemical model²³ and a two-dimensional model¹⁶. Furthermore, the model calculations discussed earlier indicate that inequality (2) is satisfied at flight altitudes throughout our latitude range. Therefore all information available supports inequality (3) for these low values of N₂O. Implicit in the above discussion is that the formulations do have limits of applicability, the most obvious one being that they do not apply at altitudes above the maximum stratospheric mixing ratios for O₃. For at such altitudes, N₂O and O₃ are both decreasing with altitude but inequality (3) implies that as N₂O decreases, O₃ increases. With this reservation a lower-limit estimate for unperturbed O₃ both outside and with increased reservation inside the boundary can now be calculated. Figure 4a, b shows measured O₃ against latitude with respect to the boundary where O₃ is averaged over all ten flights from 23 August to 22 September at both flight levels. Otherwise data are treated as in Fig. 3 for O₃/NO_y. Also plotted is the corresponding minimum value of unperturbed O₃ calculated from inequality (3). Figure 4c shows the O₃ loss with respect to the boundary, calculated as the percentage decrease in O₃ and based on Fig. 4a, b. The loss outside the boundary is typically 20–25% at 425 K for the first 5° with a maximum value indicated of ~30%. From Fig. 3 we obtained a maximum value of 30% O₃ loss outside the boundary, which is virtually identical to the loss shown in Fig. 4c, although the analysis of Fig. 3 includes only seven of the ten flights included in Fig. 4. This difference in number of flights is because particle-free NO_y data are available only from seven flights. Both analyses indicate that the loss is greater at the lower level, which is consistent with diabatic

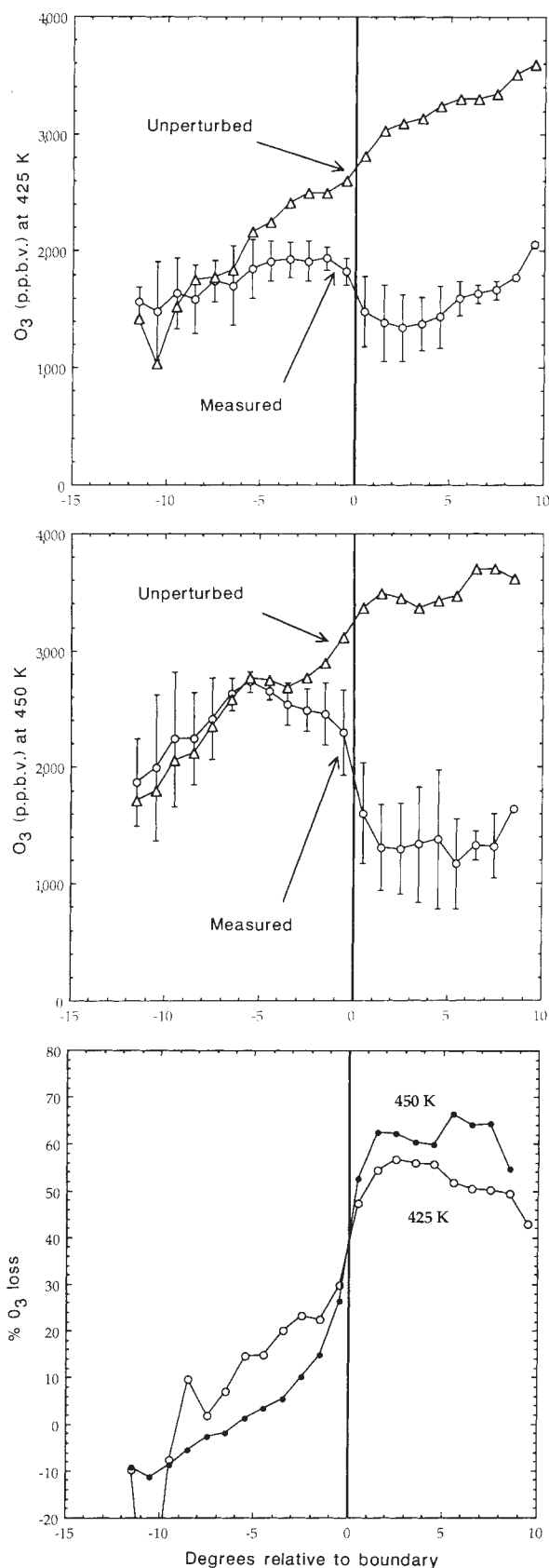


FIG. 4 a, O₃ against latitude referenced to the boundary (heavy vertical line) for the 425-K flight level. Data are averaged over intervals of 1° latitude and over all ten flights from 23 August to 22 September 1987. b, As in a except at 450-K flight level. c, Average percent O₃ loss over 30 days derived from a and b.

descent and a simultaneous chemical loss of O_3 .

This analysis can also be applied to individual flights rather than averages of flights. For example, from data taken on 23 August, the first day that a flight clearly penetrated the boundary, we calculate a 40% loss in O_3 inside the boundary. On the last mission flight date, 22 September, still two weeks before the minimum column O_3 was measured in 1987 (ref. 24), there was a 70–80% loss within the boundary in the layer between 17.5 km and 19 km. This is substantially more than the 53% isentropic temporal decrease observed in the same region from 23 August to 22 September¹⁸.

Discussion and conclusions

Since the ozone hole was first identified in 1985 it has been considered to be a recurring springtime event with a nearly monotonic increase of intensity over its history of ~10–15 years²⁵. Here we have presented strong evidence of significant O_3 loss occurring as early as mid-August both inside and outside the highly chemically perturbed region, and that this loss is chemical in origin rather than a result of atmospheric dynamics. Although the mechanism for the chemical loss outside the boundary cannot be inferred from this approach, there is evidence of a temporal column O_3 decrease over the past decade in this region, therefore indicating that the chemical loss that we report has an anthropogenic component. In particular, ref. 26 shows, by comparing satellite data from 1979–80 with 1986–87, a nine-year temporal decrease from 6% to 18% in column O_3 between 60° S and 66° S during August and September, a region near, but on the average outside, the boundary during AAOE. In ref. 27 the data were analysed with a linear fit of the yearly data, starting with 1978 and ending with 1988. This analysis produced virtually identical trends from 60° S to 66° S from mid-August through September. Both studies also show a long-term temporal decrease of more than 10%, poleward of 66° S during August. These analyses present a consistent picture of mid-August O_3 decreases, both inside and outside the vortex, that have intensified during the past ten years. The results we have presented co-locate a region of chemical O_3 loss with these long-term temporal decreases in O_3 , suggesting that at least part of the loss has an anthropogenic origin.

The cause of the O_3 loss outside the boundary, where the measured ClO is only 5–10% of the level measured inside the boundary, is not known, and more experimental investigation is warranted. But it has been suggested by Tuck¹¹ that O_3 losses may occur near the boundary and outside the highly chemically perturbed region during June, July and August as long as there is sunlight. Such losses would be due to reactive chlorine, as is the case within the ozone hole itself. Support for this conclusion comes from photochemical model studies made along isentropic trajectories outside the boundary²⁸. These studies point out that

the discrepancy between observed and measured values of ClO and NO indicates significantly perturbed photochemistry in this region without substantial denitrification. In sunlight the formation of polar stratospheric clouds (PSCs) outside the boundary can perturb chlorine and nitrogen chemistry, initiating catalytic destruction cycles. Observed temperatures are above the frost point on average, but low enough for PSCs to form that contain nitric acid, a component of NO_y ^{13,29}. The NO_y measurements indicate that PSC formation does not generally lead to denitrification. The maximum frequency of PSC formation over the continent and inside the vortex occur to the east of the ER-2 flight tracks over the Weddell Sea³⁰, downwind from the ER-2 flight track. The perturbed levels of reactive chlorine initiated by the PSCs cannot be long-lived because of the lack of denitrification. In sunlight, nitric acid will gradually photolyse to produce NO_2 , and this will reduce ClO levels to near unperturbed values. With such intermittent chemical perturbations, this region may contain sufficient recurrent enhancements in ClO to destroy O_3 to the extent indicated in our analysis.

Obtaining values for unperturbed O_3 in the Antarctic from other species must carry with it some uncertainty, but the analysis presented here has been conservative. In particular, if the value of 300 had been chosen for O_3/NO_y as is indicated by the Southern Hemisphere data, the analysis would have produced somewhat larger O_3 losses. The negative O_3 loss 10–15° outside the boundary (Figs 3 and 4c) shows that the choice of 270 for the ratio is conservative. Nevertheless, the method is more reliable than the simple 'before and after' comparisons of O_3 profiles which we strongly caution against when analysing losses of O_3 at high latitudes unless the losses are very large, as within the Antarctic ozone hole.

We have analysed winter-time high-latitude O_3 changes in a region where O_3 is considered under dynamical control and standard gas-phase photochemistry is extremely slow. We have estimated a reference state for O_3 in this region from other conserved trace species and compared this with measured values to determine a minimum loss in ozone. In this context we can summarize the important conclusions as follows: (1) There is a substantial O_3 loss by mid-August both inside and outside the highly chemically perturbed region; (2) For the first 5° outside the boundary, the average O_3 loss from 23 August to 22 September was at least 15–30% at 17.5 km and 4–25% at 19 km; (3) The O_3 loss found outside the boundary is chemical in origin and did not occur within the highly chemically perturbed region. An anthropogenic component cannot be ruled out; (4) For the first 5° inside the boundary and in the layer between 17.5 and 19 km the air observed on 23 August had an O_3 loss of about 40%. By 22 September there was an O_3 loss of at least 75%, much larger than the average isentropic temporal decrease of 53% from 23 August to 22 September.

Received 9 June; accepted 22 September 1989.

- Atkinson, R. J., Matthews, W. A., Newman, P. A. & Plumb, R. A. *Nature* **340**, 290–294 (1989).
- Farman, J. C., Gardiner, B. G. & Shanklin, J. D. *Nature* **315**, 207–210 (1985).
- Stolarski, R. S. *et al.* *Nature* **322**, 808–811 (1986).
- Hofmann, D. J., Harder, J. W., Rolf, S. R. & Rosen, J. R. *Nature* **326**, 59–62 (1987).
- Komhyr, W. D., Oltmans, S. J. & Grass, R. D. *J. geophys. Res.* **93**, 5167–5184 (1988).
- Solomon, S., Garcia, R. R. & Stordal, F. *J. geophys. Res.* **90**, 12981–12989 (1985).
- Garcia, R. R. & Solomon, S. *J. geophys. Res.* **90**, 3850–3868 (1985).
- Proffitt, M. H. *et al.* *J. geophys. Res.* **94**, 11437–11448 (1989).
- Danielsen, E. F. & Houben, H. *Anthropogene Beeinflussung der Ozonschicht*, 191–242 (DEHEMA, Frankfurt am Main, 1988).
- Proffitt, M. H. *et al.* *J. geophys. Res.* (in the press).
- Tuck, A. F. *J. geophys. Res.* **94**, 11687–11737 (1989).
- Kelly, K. K. *et al.* *J. geophys. Res.* **94**, 11317–11357 (1989).
- Fahey, D. W. *et al.* *J. geophys. Res.* (in the press).
- Galbally, I. E. *et al.* *Q. J. R. met. Soc.* **112**, 775–809 (1986).
- Ko, M. K. W., McElroy, M. B., Weisenstein, D. K. & Sze, N. D. *J. geophys. Res.* **91**, 5395–5404 (1986).
- Solomon, S. & Garcia, R. R. *J. geophys. Res.* **89**, 11633–11644 (1984).
- Russell, J. M. III *et al.* *J. geophys. Res.* **93**, 1718–1736 (1988).

- Proffitt, M. H. *et al.* *J. geophys. Res.* (in the press).
- Starr, W. L. & Vedder, J. F. *J. geophys. Res.* **94**, 11449–11463 (1989).
- Margitan, J. J. *et al.* *J. geophys. Res.* (in the press).
- Loewenstein, M., Podolske, J. R., Chan, K. R. & Strahan, S. E. *J. geophys. Res.* **94**, 11589–11598 (1989).
- Podolske, J. R., Loewenstein, M., Strahan, S. E. & Chan, K. R. *J. geophys. Res.* (in the press).
- Brasseur, G. & Solomon, S. *Aeronomy of the Middle Atmosphere*, 2nd edn (Reidel, Dordrecht, 1986).
- Krueger, A. J. *et al.* *The 1987 Airborne Antarctic Ozone Experiment: The Nimbus-7 TOMS Data Atlas* (NASA Ref. Publ., NASA RF-1201, Goddard Space Flight Center, 1988).
- Solomon, S. *Rev. Geophys.* **26**, 131–148 (1988).
- International Ozone Trends Panel Report: 1988 (WMO Report No. 18, in the press).
- UNEP/WMO Stratospheric Assessment of Stratospheric Ozone: 1989 (UNEP/WMO Report, in the press).
- Jones, R. L. *et al.* *J. geophys. Res.* **94**, 11529–11558 (1989).
- Chan, K. R., Scott, S. G., Bui, T. P., Bowen, S. W. & Day, J. *J. geophys. Res.* **94**, 11573–11587 (1989).
- Watterson, I. G. & Tuck, A. F. *J. geophys. Res.* (in the press).

ACKNOWLEDGEMENTS. We thank J. Margitan, S. Solomon and R. Jones for their comments. We thank The Chemical Manufacturers Association and NASA for funding.

Re-evaluation of the linkage relationship between chromosome 11p loci and the gene for bipolar affective disorder in the Old Order Amish

John R. Kelsoe*, Edward I. Ginns*, Janice A. Egeland†, Daniela S. Gerhard‡, Alisa M. Goldstein§, Sherri J. Bale§, David L. Pauls¶, Robert T. Long*, Kenneth K. Kidd¶, Giovanni Conte*, David E. Housman# & Steven M. Paul*††

* Sections on Molecular Neurogenetics and Molecular Pharmacology, Clinical Neuroscience Branch, National Institute of Mental Health, Bethesda, Maryland 20892, USA

† Department of Psychiatry, University of Miami School of Medicine, Miami, Florida 33101, USA

‡ Department of Genetics, Washington University School of Medicine, St Louis, Missouri 63110, USA

§ Family Studies Section, Environmental Epidemiology Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

¶ Child Study Center and Department of Human Genetics, Yale University School of Medicine, New Haven, Connecticut 06510, USA

¶ Department of Human Genetics and Psychiatry, Yale University School of Medicine, New Haven, Connecticut 06510, USA

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Reanalysis of an Old Order Amish pedigree, to include several new individuals and two changes in clinical status, markedly reduces the probability of linkage between bipolar affective disorder and the Harvey-*ras-1* oncogene and insulin loci on chromosome 11. This linkage can be excluded using a large lateral extension of the original Amish pedigree.

THE affective disorders are a common¹ and heterogeneous group of illnesses that are characterized by abnormalities of mood and cognition, as well as neurovegetative functions such as sleep, appetite and libido². Although the aetiology of these disorders is unknown, genetic factors appear to contribute to the development of both bipolar and unipolar depression³ in many individuals whose families have multiple affected members^{4,5}. Attempts to delineate genetic loci for the various affective disorders have been hampered by the likely presence of genetic heterogeneity as well as nongenetic forms (phenocopies) of the disorder. Several strategies have evolved to circumvent these problems. Perhaps the most powerful of these has been the application of genetic linkage techniques to the study of populations that have developed from a restricted gene pool and hence have a more restricted array of alleles predisposing to affective disorder. Egeland and co-workers⁶ have identified several large pedigrees with a high incidence of affective disorder among the Old Order Amish located in southeastern Pennsylvania. The Old Order Amish are a religious sect numbering ~15,000 who descend from some 30 pioneer couples and who have remained genetically isolated, thereby minimizing the introduction of multiple genes responsible for inherited disorders. Amish families have large sibships and multiple living generations, making them ideal for genetic studies. Alcohol and drug abuse, which often complicate psychiatric diagnoses, are rare among the Amish. Bipolar affective disorder, however, occurs with a prevalence rate, characteristic symptom pattern and clinical course that are similar to those in the general North American population⁷. The identification and characterization of these pedigrees^{6,7} led to the initiation of early genetic linkage studies

but no evidence for linkage between various polymorphic serum proteins or blood group antigen loci and affective disorder was found⁸. More recently, using a molecular genetic approach, Egeland and colleagues reported evidence supporting the localization of a gene conferring a strong predisposition to bipolar affective disorder linked to two loci located on the short arm of chromosome 11, the Harvey-*ras-1* oncogene locus (*HRAS*) and the insulin (*INS*) locus⁹. Several studies in non-Amish families have excluded linkage between a single locus predisposing to affective disorder and these chromosome 11 markers^{10,11}.

The current set of analyses extend the original study by adding genotypic data obtained on several members not previously included, by incorporating new clinical data on several recent onsets of affective illness in previously unaffected members of the pedigree, and by extending the number of individuals under study with a large lateral extension of the original pedigree. The addition of this new clinical and genetic data identified individuals in the core pedigree who were obligate recombinants with either marker when the diagnostic criteria used in the original study were applied. The impact of these new data was to lower the overall lod score and hence weaken the evidence for linkage between an affective disorder locus on 11p15 and the *HRAS* and *INS* loci. Also, the large lateral extension of pedigree 110 does not show linkage between these chromosome 11p15 markers and an affective disorder locus. When combined with data from the original core pedigree, the hypothesis that a single affective disorder locus is linked to *HRAS* and *INS* in the entire pedigree could be excluded at a distance of 5 centimorgans (cM; a centimorgan is equivalent to 1% recombination between two loci and corresponds to $\sim 1 \times 10^6$ base pairs of DNA) for *INS* and 10 cM for *HRAS*. We conclude that the evidence for linkage between bipolar affective disorder and the *HRAS* and *INS* genes in the Old Order Amish is not as strong as was initially suggested.

The extended and updated pedigree

The original pedigree 110, as previously reported, consisted of 81 subjects and includes the extended families of 6 ill probands: 5 with bipolar affective disorder and 1 with schizoaffective disorder, circular type; as well as their first- and second-degree relatives (Fig. 1). These patients were assessed by family study methods¹². They and their relatives were interviewed using the Schedule for Affective Disorder and Schizophrenia-Lifetime

†† To whom correspondence should be addressed.

Version (SADS-L)¹³. In addition, all hospital and clinic records were abstracted. The multiple SADS-L interviews and collated records were evaluated by a five-member psychiatric review board who were blind to patient identity and biological relationships. Most patients were diagnosed by the Board at two points in time and the reliability of diagnoses between assessors was high¹². Nineteen of the 81 subjects in the original pedigree 110 were given diagnoses of major affective disorder according to Research Diagnostic Criteria (RDC)¹⁴. Fourteen had some form of bipolar disorder and five had unipolar depression. Five persons had minor depressive disorder by RDC and were treated as 'unaffected' in all the linkage analyses. One person who failed to meet RDC diagnostic criteria was diagnosed as having a 'bipolar-type' disorder by consensus of the psychiatric panel. Sixteen of 19 individuals with major affective disorder have been hospitalized at least once, yielding a total of 86 abstracted hospital records for diagnostic review. The availability of these multiple psychiatric evaluations contributed to the reliability of diagnoses in this study. The updated core pedigree includes data on two new unaffected individuals, new genotypes on ten unaffected members not included in the original study, and changes in diagnoses for two individuals owing to onset of illness subsequent to the initial linkage analysis.

The core pedigree of 81 subjects has been extended to include a small left extension closely related to the core pedigree through common great-grandparents (Fig. 1). This left extension consisting of six subjects (two with diagnoses of bipolar affective disorder) has been included in the linkage analysis of the updated core pedigree (Fig. 1 and Table 2). In addition, a large 31-member right extension (Fig. 1) has been studied; it contains

eight individuals with a diagnosis of major affective disorder (four with bipolar disorder and four with major depressive disorder). Inspection of Fig. 1 reveals that individual C-4 connects this right extension to the core pedigree. This person and her spouse were psychiatrically well. Both, however, had uncles with RDC diagnoses of bipolar affective disorder (type I) by psychiatric board review. The decision to extend the pedigree through individual C-4, rather than C-3, was based on sibship structure and the segregation of the illness over four generations. Historical and genealogical records of the upper generation (generation B, Fig. 1) of this right extension suggests the possibility that affective disorder in this extension may have been introduced by an Amish-Mennonite family whose surname is no longer found among members of the Amish community (Fig. 1). Consequently, for the linkage analysis, the right extension was analysed both separately and combined with the updated core pedigree (Table 2).

As part of the Amish Study, each subject (for whom immortalized cell lines have been established) is visited annually by a staff member. Field work is carried out by staff members who are blind to the genotypes of all subjects. As a result of this longitudinal evaluation, it has been noted that since the initial analysis of the data, the clinical status of two individuals (III-21, IV-20) (Fig. 1) has changed from unaffected to affected.

Defining a diagnostic hierarchy

Relatives of individuals with bipolar affective disorder have been found to have a variety of affective and nonaffective psychiatric syndromes¹⁵. These include individuals with

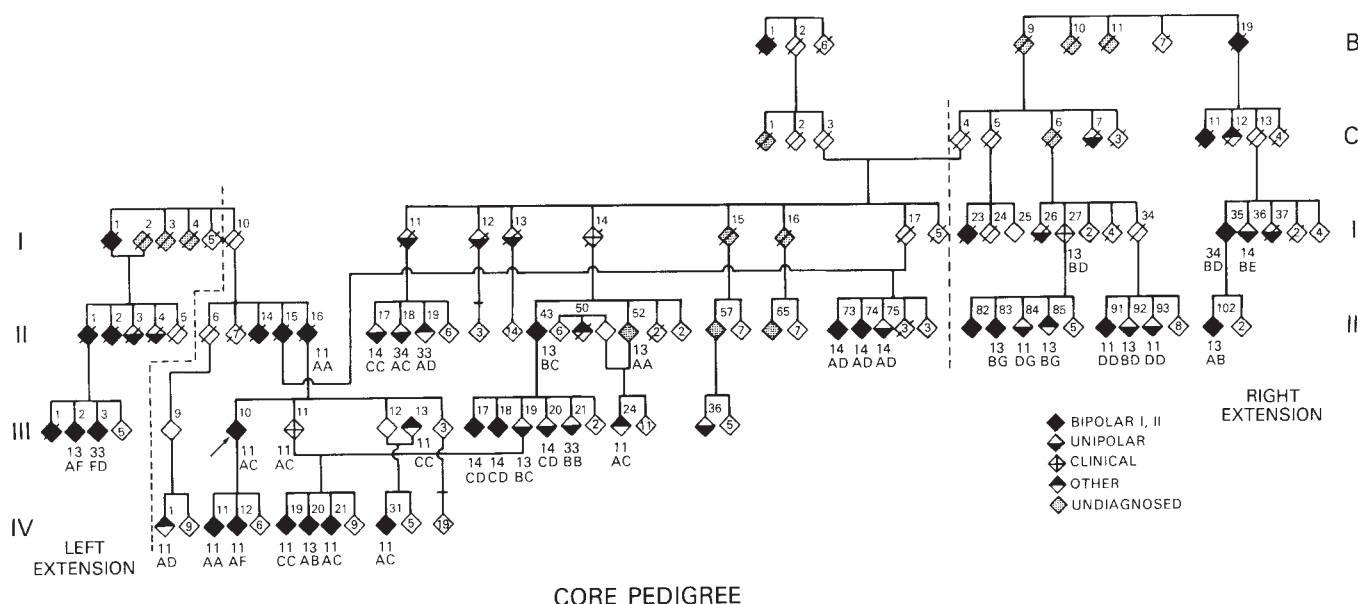


FIG. 1 Pedigree 110 and its extensions. The structure of the 120 member family (pedigree 110) used in this study is shown. The dotted lines distinguish the left and right extensions from the core pedigree. Sex and birth order have been disguised to protect anonymity. Genotypes are not included for the unaffected members to protect those for whom a genetic vulnerability might be indicated. Of those with diagnoses designated 'other', subject III-24 has schizoaffective illness, the rest have minor depressions. Genotypes are designated with numbers, 1-4, for *HRAS*, and with letters, A-G, for *INS*. In several members of the right extension, one laboratory (NIMH) identified another *INS* allele, H, which is 2.4 kb in size, and, therefore, difficult to distinguish from the D allele. In the analyses reported here, the H allele was scored as D, because of the difficulty in resolving the two alleles. The linkage analyses also were carried out using the H allele resulting in somewhat more negative lod scores, but changing neither the basic conclusions nor the size of the exclusionary range (data not shown). Genotypes were determined independently in two laboratories, NIMH and St Louis. In the initial analysis several discrepancies in genotyping were obtained and were resolved by reanalysis in both laboratories except for one individual (III-20),

and a new DNA sample will be needed to resolve this discrepancy. The St Louis genotypings were carried out as previously described⁹. In the NIMH protocol, genomic DNA was digested with the appropriate endonuclease (*Sst*I for *HRAS* or *Pvu*II for *INS*; BRL) and 10 µg was electrophoretically separated in either 0.8% agarose and Oncor buffer (*HRAS*), 1.0% agarose and TBE (insulin alleles A, B, C, E, F), or 1.0% agarose and Oncor buffer (insulin alleles D, G). The DNA was transferred by vacuum to nylon membranes (Oncor). Probes for *HRAS* (pEJ 6.6) and *INS* (phins 310) were obtained from American Type Culture Collection (Rockville) and labelled by the random primer method²² using [α -³²P]dCTP (Amersham). Hybridization was carried out at 45 °C overnight using the appropriate probe and Hybrisol I (Oncor) with poly-A RNA (10 µg ml⁻¹) and denatured *E. coli* DNA (100 µg ml⁻¹) added. Membranes were washed three times in 0.1% SDS, 0.1 × SSC at room temperature for 15 min, followed by one wash at 52 °C for 1 h. Autoradiography was performed using Kodak XAR film and enhancing screens. Films were read independently by two people unaware of the diagnoses.

schizoaffective disorder, unipolar depressive disorder, cyclothymia, dysthymia or depressive neurosis and various personality disorders with and without affective symptoms. As pedigree 110 and its extensions include a number of individuals with these psychiatric diagnoses (Fig. 1), which may (or may not) be genetically related to bipolar affective disorder, a diagnostic hierarchy was developed to include such individuals in the linkage analysis. This hierarchical system has two dimensions. The first measures severity of affective symptoms and has four categories (Table 3). The other dimension includes milder syndromes in the affected phenotype by allowing for both a criteria-based system such as the RDC (as was used in the initial study; ref. 9) as well as the clinical impressions of psychiatrists on the review board. The diagnostic hierarchy used in the present study (Table 3 legend) was developed by members of the psychiatric review board who were unaware of the genotypes of any of the members of the pedigree.

Genotyping and linkage analysis

Genotyping was performed independently in two laboratories (NIMH and St Louis). Genomic DNA was extracted from lymphoblastoid cell lines obtained from the Coriell Institute for Medical Research¹⁶ for 120 members of Amish pedigree 110 (Fig. 1 and Table 2). In one laboratory (St Louis), DNA for some family members was extracted directly from blood. The methodology for DNA extraction and Southern analysis differed slightly between the two laboratories (see legend Fig. 1). A

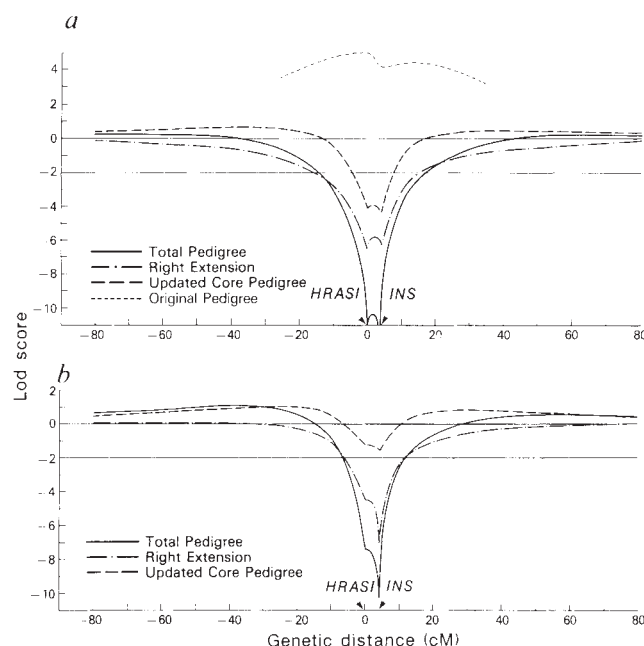


FIG. 2 Multipoint linkage analysis using *A*, the original diagnostic criteria and *B*, the diagnostic system which includes broader diagnostic criteria to define the affected phenotype. System *A* is equivalent to system number 3 (RDC) from Table 3, and includes affected individuals with bipolar disorder, major affective disorder and schizoaffective disorder, but not minor depression. System *B* is identical to system 2 using clinical criteria from Table 2. It includes as affected those with bipolar, schizoaffective or recurrent unipolar major affective disorders, but excludes those with minor depression or only a single episode of major depression. For each diagnostic system, the data has been analysed for the updated core pedigree, the right extension and the total pedigree. Three-point analyses to determine the most likely location of the locus for bipolar affective disorder in relation to *HRAS* and *INS* were performed using the LINKMAP subroutine of LINKAGE version 4.8 (ref. 20). For these analyses, the distance between *HRAS* and *INS* was held constant at 3.8 cM (ref. 23). Five liability classes were defined (see Table 1). The recombination distance between *HRAS* and *INS* was examined in each portion of the pedigree and found to be consistent with other estimates²³.

TABLE 1 Age-specific penetrances used for multipoint linkage analyses

Age	Liability class	Probability of disease		
		AA	Aa	aa
0-14	1	0.0001	0.0001	0
15-19	2	0.18	0.18	0.0001
20-24	3	0.42	0.42	0.001
25-29	4	0.73	0.73	0.001
≥ 30	5	0.85	0.85	0.001

Values represent age- and genotype-specific penetrances of the Major Affective Disorder (MAD) phenotype. AA represents an individual homozygous for the disease alleles; Aa represents heterozygotes; and aa represents individuals genotypically normal at the MAD locus.

consensus was obtained on all typings, however, with the exception of one individual (III-20). Resolution of this single discrepant genotype will require obtaining a new DNA sample on individual III-20 as each laboratory was able to confirm their typing on several occasions from their original source of DNA, indicating that an error had occurred in the handling and (or) storage of lymphocytes for this subject. Nevertheless, the linkage analyses were performed twice using each genotyping for this subject without any effect on the lod scores (data not shown).

Four *HRAS* alleles (about 5.2 kilobases (kb) long (1); ~5.7 kb (2); ~6.2 kb (3); ~6.8 kb (4)) and four *INS* alleles (about 760 bp (A); ~800 bp (B); ~830 bp (C); ~2.35 kb (D)) were delineated as described in the original study⁹. Three additional *INS* alleles (~650 bp (E); ~820 bp (F); ~2.5 kb (G)) were also observed, primarily in the extensions of the core pedigree. The microheterogeneity observed at the *INS* locus is consistent with other population studies of this gene¹⁷.

Analyses of the cosegregation of alleles at the *HRAS* and *INS* loci with the bipolar affective disorder phenotype were performed using the computer programs LIPED^{18,19} and LINKAGE²⁰ (see Table 2 legend). Analysis of linkage between affective disorder and each marker was done for the original core pedigree, the updated core pedigree, the right extension and the complete pedigree. Five liability classes were defined and the age-specific penetrances for the multipoint linkage analysis are shown in Table 1.

Linkage in the updated core pedigree

In the course of our study we (J.K., E.G., S.B., A.G., R.L., G.C. and S.P., unpublished data) independently reanalysed the original core pedigree with the affection status and diagnoses used in the initial study, but with the genotypes from the NIMH laboratory, and identical lod scores were obtained. The genotypes of 2 new unaffected family members (II-20, 21) added into the core pedigree had only a minor effect on the lod scores. The addition of genotypes for 10 unaffected individuals, however, who were incompletely genotyped in the original analysis (data from only one of the two marker loci were available), resulted in a significant change in the lod scores for both *HRAS* and *INS*. The maximum lod score for *HRAS* decreased from 4.08 at $\Theta=0$ to 2.46 at $\Theta=0.05$, whereas that for *INS* increased from 2.63 to 3.38 at $\Theta=0$ (Table 2).

The most dramatic effect on the lod scores resulted from the addition of updated diagnostic information on two family members who became ill after the original study; one (IV-20) developed a manic psychosis (bipolar affective disorder, type 1) and the other (III-21) suffered a severe depression with catatonic features (major depressive disorder). Together, these changes lowered the interpolated maximum lod scores to 1.03 at $\Theta=0.17$ and 1.75 at $\Theta=0.14$ for *HRAS* and *INS*, respectively (Table 2). The addition of the left extension to the core pedigree (updated core) further reduced the lod scores to a maximum of 0.70 at $\Theta=0.20$ and 1.29 at $\Theta=0.20$ for *HRAS* and *INS*, respectively (Table 2).

TABLE 2 Lod scores resulting from sequential expansion of the core pedigree

Information added	Recombination frequency (Θ)					
	0.00	0.001	0.05	0.10	0.20	0.30
Locus: <i>HRAS</i>						
1. Original core ($N=81$)	4.08	4.08	3.91	3.63	2.85	1.81
2. 1 + new genotypes ($N=83$)	2.36	2.37	2.46	2.39	1.92	1.25
3. 2 + updated DX	-1.75	-1.53	0.44	0.91	1.01	0.63
4. Updated core (3 + left ext., $N=89$)	-3.22	-2.97	-0.34	0.36	0.70	0.46
5. Right ext. ($N=31$)	-6.15	-5.83	-2.95	-1.93	-0.85	-0.32
6. Total (4 + 5, $N=120$)	-9.31	-8.93	-3.97	-2.21	-0.65	-0.21
Locus: <i>INS</i>						
1. Original core ($N=81$)	2.63	2.63	2.50	2.31	1.85	1.50
2. 1 + new genotypes ($N=83$)	3.38	3.38	3.24	3.02	2.43	2.23
3. 2 + updated DX	-0.19	0.05	1.45	1.69	1.62	1.21
4. Updated core (3 + left ext., $N=89$)	-1.90	-1.64	0.41	0.99	1.29	1.07
5. Right ext. ($N=31$)	-4.81	-4.54	-2.46	-1.73	-0.84	-0.35
6. Total (4 + 5, $N=120$)	-7.75	-6.71	-2.40	-1.07	0.19	0.55

The original core pedigree is identical in membership, genotypes and diagnoses to that reported originally⁹. Each subsequent analysis includes the information in the previous one plus the additional data as indicated. The new genotypes include 10 *HRAS* and *INS* genotypes which were not available at the time of the original study. The updated diagnoses refer to two individuals who had an onset of major affective disorder subsequent to the initial analysis (IV-20, bipolar I disorder; II-21, major depressive disorder). Analyses of the cosegregation of alleles at the *HRAS* and *INS* loci with the bipolar affective disorder phenotype were performed using the computer programs LIPED^{18,19} and LINKAGE²⁰ on a COMPAQ Deskpro 386 microcomputer and a VAX 8800. All pairwise analyses (*HRAS* $\times$ *INS*, *HRAS* $\times$ affective disorder, *INS* $\times$ affective disorder) used LIPED. Analyses were performed assuming an autosomal dominant transmission model. For the two-point analyses, the probability of exhibiting the affected phenotype for individuals who inherited the predisposition allele for bipolar affective disorder was modelled as a linear function of age with a penetrance of 0 at age 15 and a maximum penetrance of 85% reached at 35 years of age. The maximum penetrance for the nonsusceptible genotype was 0.1%, reached at 35 years of age. The allele frequency of the disease-predisposition allele was taken as 0.021 for all analyses. The *Sst*I *HRAS* allele frequencies were 0.30, 0.30, 0.10, and 0.30 for alleles 1, 2, 3, and 4, respectively. Genotypes at the *INS* locus were reduced to a four-allele system to maximize efficiency of computing time. The latter involves no loss of information in the analyses. All alleles were considered as equally likely.

Linkage in the right extension

When analysed alone, the lod scores for linkage between either the *HRAS* or *INS* loci and bipolar affective disorder in the right extension were strongly negative (Table 2); linkage can be excluded for $\Theta \leq 0.05$. Similarly, when the right extension and core are considered together, linkage can be excluded for $\Theta \leq 0.10$ for *HRAS* and $\Theta \leq 0.05$ for *INS*. The lod scores at $\Theta = 0$ represent odds of 10^{10} and 10^8 against close linkage between a disease locus and the *HRAS* and *INS* loci, respectively.

Alternative diagnostic criteria

Linkage analyses were done using alternative diagnostic criteria (clinical consensus diagnoses) as outlined in Table 3 on the core pedigree. The strongest evidence for linkage was obtained when all individuals with mania, hypomania or recurrent major depression were considered affected and when clinical, as opposed to RDC, diagnoses are used (maximum lod score for *INS* of 2.88 at $\Theta = 0$) (Table 3). Inclusion of individuals with minor depression decreased the lod scores for both *HRAS* and *INS*.

Multipoint linkage analysis

Illustrated in Fig. 2a are the results of the multipoint linkage analysis between *HRAS*, *INS* and a locus for bipolar affective disorder using RDC diagnostic criteria. Linkage was excluded in the interval between *HRAS* and *INS*, as well as to within 5 cM on either side of each of these markers for the updated core pedigree. Both the right extension alone and the total pedigree gave exclusions to within 15 cM on either side of the *HRAS*-*INS* interval.

Figure 2b illustrates the multipoint linkage results when a clinical (as opposed to RDC) diagnostic system is used to define the affected phenotype including mania, hypomania and recurrent major depressive disorder. Using this diagnostic system, linkage could not be excluded in the core pedigree alone; the maximum lod score was ~ 1.0 at 26 cM distal to *HRAS*. Linkage could be excluded in the right extension and in the total pedigree within the interval between *HRAS* and *INS* and to within 7 cM on either side of the markers.

Discussion

The initial linkage analyses of pedigree 110 reported by Egeland *et al.*⁹ provided strong evidence for a locus conferring a predisposition to manic-depressive illness on chromosome 11p15. These investigators reported a lod score of 4.08 at $\Theta = 0$ in the two-point analyses between *HRAS* and bipolar affective disorder. The multipoint linkage analysis using both markers, *HRAS* and *INS*, yielded a lod score of ~ 5 (odds of 10^5 to 1 in favour of linkage). In our reanalysis, the recent onset of affective disorder in 2 members of the pedigree and the addition of genotypic data on six individuals in a newly included sibship, and on ten previously incompletely genotyped members, markedly reduces the statistical probability of linkage in both the two-point and multipoint analysis (Table 2 and Fig. 2).

Linkage between bipolar affective disorder and the *HRAS* and *INS* loci can also be excluded in the large lateral extension of pedigree 110. This observation can be interpreted in at least two ways: that there is genetic heterogeneity among the Old Order Amish; and/or that, if it is assumed that the illness in the original core and this extension is genetically homogeneous, there is strong evidence against linkage of a susceptibility locus for bipolar affective disorder to this region of 11p15. Although the Old Order Amish represent a genetic isolate, it has been suggested (J. A. Egeland, personal communication) that multiple genes may be responsible for affective disorder within the Amish population as several progenitor couples had mental illness among their first and second generation descendants. The analysis, therefore, of a large portion of the core pedigree may be complicated by the presence of two different genes responsible for affective illness.

What then is the evidence that bipolar affective disorder is linked to the *HRAS* and *INS* loci in the Old Order Amish? The hypothesis that has been tested so far is that the inheritance of specific alleles at a locus on chromosome 11p15 linked to *HRAS* and *INS* is a necessary condition for the occurrence of affective disorder in the members of pedigree 110. This hypothesis has been rejected in the analysis of the extended pedigree. Alternative hypotheses consistent with the expanded data set and analysis include the possibility that linkage to 11p15 markers in the

TABLE 3 Lod scores in core pedigree using alternate diagnostic criteria

Diagnoses included	Recombination frequency (Θ)						
	0.00	0.05	0.10	0.15	0.20	0.25	0.30
Locus: <i>HRAS</i>							
1. Mania or hypomania only							
RDC	0.29	0.65	0.78	0.81	0.78	0.69	0.57
Clinical	0.53	0.59	0.61	0.60	0.56	0.49	0.40
2. 1 + recurrent MDD							
RDC	0.69	1.17	1.33	1.32	1.19	0.99	0.73
Clinical	1.34	1.49	1.49	1.39	1.21	0.97	0.69
3. 2 + MDD single episode							
RDC	-1.02	0.71	1.06	1.15	1.08	0.92	0.68
Clinical	-0.59	1.02	1.23	1.22	1.10	0.90	0.65
4. 3 + minor dep.							
RDC	-4.23	-1.30	-0.53	-0.13	0.08	0.17	0.17
Clinical	-3.66	-0.99	-0.37	-0.05	0.10	0.16	0.14
Locus: <i>INS</i>							
1. Mania or hypomania only							
RDC	1.03	1.09	1.08	1.02	0.92	0.78	0.63
Clinical	1.28	1.19	1.07	0.94	0.81	0.68	0.54
2. 1 + recurrent MDD							
RDC	2.22	2.30	2.25	2.11	1.91	1.65	1.35
Clinical	2.88	2.80	2.63	2.39	2.10	1.77	1.42
3. 2 + MDD single episode							
RDC	0.36	1.82	1.98	1.94	1.80	1.58	1.30
Clinical	0.94	2.33	2.36	2.22	1.99	1.70	1.37
4. 3 + minor dep.							
RDC	-3.35	-0.14	0.50	0.79	0.90	0.90	0.80
Clinical	-2.76	0.36	0.88	1.07	1.10	1.02	0.87

The effect of different diagnostic criteria on the lod score analysis in the updated core pedigree. These data were calculated from a version of the updated core pedigree that lacks the small left extension, but includes all other updated information. See the legend of Table 2 for details of the two-point linkage analyses. Various diagnostic criteria (RDC, clinical) were used by the psychiatric review board who were blind to genotypes or the effects of diagnoses on the lod scores. The hierarchy is devised along two dimensions: one is based on the degree of relatedness of a given diagnosis to bipolar affective disorder, the other is based on whether the diagnoses were made by RDC criteria or by the clinical impressions of the review board. For each diagnostic system, lod scores are reported for both RDC and clinical diagnoses. The first diagnostic system includes as affected only those individuals who have had mania or hypomania. The next adds those individuals who have had multiple episodes of major depressive disorder (MDD), followed by a system that includes those who have had only a single episode of MDD. The final diagnostic system includes any individual who has had a diagnosis of minor depressive disorder, which requires fewer symptoms than major depression.

core is due merely to chance. Many other DNA markers have been tested in the core pedigree but no evidence for linkage has yet been found²¹. Testing of markers spanning the remainder of the genome will lead to the identification of the site of this locus, if a single major locus hypothesis is correct. Even a complete systematic evaluation of pedigree 110, however, using highly polymorphic markers spaced very closely and covering the entire genome may not resolve the issue. If genes at more than one locus can independently lead to genetic susceptibility to affective disorder in this pedigree, then conventional linkage analysis over the entire pedigree will continue to be problematic. It may, therefore, be impossible to generate a strong positive lod score over the entire pedigree for linkage of a single major affective disorder locus to any marker.

Another important consideration is the possibility that non-genetic factors may contribute to the aetiology of affective disorder in pedigree 110. Any individual in whom illness was significantly mediated by environmental rather than genetic factors has an even chance of appearing as a recombinant in a linkage analysis. If the frequency of such phenocopies is low, then their impact on statistical tests of genetic hypotheses may not be great, unless they happen to occur in a critical part of the pedigree. As the frequency of such individuals increases, however, their impact on the tests of genetic hypotheses may become increasingly significant. This may be particularly problematic when diagnostic categories are broad and the population frequency of individuals considered affected becomes correspondingly high. The use of narrower or alternative diagnostic criteria may reduce the frequency of phenocopies and hence the probability of observing chance recombination. Until a systematic analysis of the entire genome has been completed,

it will not be possible to determine the extent to which these problems confound the identification of loci which contribute to the aetiology of affective disorder in pedigree 110.

Our analyses highlight many of the problems that can be anticipated in genetic linkage studies of common and complex neuropsychiatric disorders. Foremost among these are the probable presence of genetic heterogeneity even in relatively isolated populations and the presence of phenocopies, especially in large pedigrees such as in the Amish. Longitudinal follow up and reanalysis of the data incorporating unaffected members of the pedigree who subsequently become ill are essential, given the variable age of onset for psychiatric disorders. Molecular genetic studies of psychiatric disorders will continue to encounter these problems and efforts to reduce their impact will be critical. □

Received 17 July; accepted 11 October 1989.

- Reiger, D. A. et al. *Arch. Gen. Psychiat.* **45**, 977-986 (1988).
- Lehmann, H. in *Comprehensive Textbook of Psychiatry IV* (eds Kaplan, H. I. & Saddock, B. J.) 786-811 (Williams and Wilkins, Baltimore, 1985).
- Perris, C. *Acta Psychiat. Neurol. Scand. Suppl.* **42**, suppl. 194 (1966).
- Nurnberger, J. I. & Gershon, E. S. in *Neurobiology of Mood Disorders* (eds Post, R. M. & Ballenger, J. C.) 76-101 (Williams and Wilkins, Baltimore, 1984).
- Goldin, L. R., Gershon, E. S., Targum, S. D., Sparkes, R. S. & McGinnis, M. *Am. J. hum. Genet.* **35**, 274-287 (1983).
- Egeland, J. A. & Hostetter, A. M. *Am. J. Psychiat.* **140**, 56-61 (1983).
- Egeland, J. A., Hostetter, A. M. & Eshleman, S. K. *Am. J. Psychiat.* **140**, 67-71 (1983).
- Kidd, K. K. et al. *Am. J. Psychiat.* **141**, 1042-1048 (1984).
- Egeland, J. A. et al. *Nature* **325**, 783-787 (1987).
- Hodgkinson, S. et al. *Nature* **325**, 805-806 (1987).
- Detera-Wadleigh, S. et al. *Nature* **325**, 806-808 (1987).
- Hostetter, A. M., Egeland, J. A. & Endicott, J. *Am. J. Psychiat.* **140**, 62-66 (1983).
- Endicott, J. & Spitzer, R. L. *Arch. Gen. Psychiat.* **35**, 837-844 (1978).
- Spitzer, R. L., Endicott, J. & Robins, E. *Arch. Gen. Psychiat.* **35**, 773-782 (1978).

15. Nurnberger, J. I. & Gershon, E. S. in *Handbook of Affective Disorders* (ed. Paykel, E. S.) 126–145 (Guilford, New York, 1982).
16. Egeland, J. A., in *Catalog of Cell Lines: NIH Publication 87-2011* 298–306 (National Institutes of Health, Washington, D.C., 1987).
17. Rotwein, P., Yokoyama, S., Didier, D. K. & Chirgwin, J. M. *Am. J. hum. Genet.* **39**, 291–299 (1986).
18. Ott, J. *Am. J. hum. Genet.* **28**, 528–529 (1976).
19. Hodge, S. E., Morton, L. A., Tideman, S., Kidd, K. K. & Spence, M. A. *Am. J. hum. Genet.* **31**, 761–762 (1979).
20. Lathrop, G. M., Lalouel, J. M., Julier, C. & Ott, J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3443–3446 (1984).
21. Kidd, J. R. *et al. J. Psychiat. Res.* **21**, 577–580 (1987).

22. Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **132**, 6–13 (1983).
23. White, R. *et al. Nature* **313**, 101–105 (1985).

ACKNOWLEDGEMENTS. We thank Cleona C. R. Allen, and members of the Amish Study Psychiatric Board: Drs Abram M. Hostetter, James N. Sussex, Jean Endicott, David R. Offord and John J. Schwab, as well as Ellen Hamilton, Trudy Welsh and Kay Kuhns for typing of the manuscript. S.M.P. thanks Drs Lewis L. Judd and Frederick K. Goodwin for their support. We also thank the Coriell Institute for Medical Research for establishing and maintaining the Amish cell-line collection. D.S.G. acknowledges the technical assistance of Steven D. Bland and was supported by the NIH. The technical support of Barbara Stubblefield (NIMH) is acknowledged.

Scanning from an independently specified branch point defines the 3' splice site of mammalian introns

Christopher W. J. Smith, Elena B. Porro, James G. Patton & Bernardo Nadal-Ginard

Laboratory of Molecular and Cellular Cardiology, Howard Hughes Medical Institute, Department of Cardiology, Children's Hospital, and Departments of Pediatrics and Cellular and Molecular Physiology, Harvard Medical School, Boston, Massachusetts 02115, USA

During pre-messenger RNA splicing the lariat branch point in mammalian introns is specified independently of the 3' splice site by the sequence surrounding the branch point and by an adjacent downstream polypyrimidine tract. The 3' splice site is dispensable for spliceosome assembly and cleavage at the 5' splice site, and is itself determined by a scanning process that recognizes the first AG located 3' of the branch point/polypyrimidine tract, irrespective of distance or sequence environment.

THE splicing of pre-messenger RNA involves the accurate and efficient removal of introns to produce mature mRNA which is a fraction of the size of the primary transcript. The splicing reaction occurs by means of two consecutive transesterification steps¹. The RNA is first cleaved at the 5' splice site, and a 2'–5' phosphodiester bond is formed between the 5' end of the intron and the 2'-OH group of a specific 'branch point' nucleotide within the intron. In the second step, the 3' splice site is cleaved, releasing the branched 'lariat' form of the intron, and the two exons are ligated together. The *cis* information for splicing is restricted to limited regions of the transcript which interact with multiple RNA and protein factors to form a 'spliceosome' complex, in which the splicing reaction takes place^{1–4}. The GU and AG dinucleotides at the 5' and 3' intron boundaries are nearly invariant in higher eukaryotes, as is the A at the branch point^{1,5–7}. In addition, there is a consensus sequence around the 5' splice site, ^cAG-GURAGU (refs 6, 7), which is involved with base pairing to U1 small nuclear RNA (snRNA)⁸, a stretch of pyrimidines typically located immediately upstream of the 3' splice site and a weak mammalian consensus branch-point sequence, YNYURAY, which binds to U2 small nuclear ribonucleoproteins⁹ (snRNPs) potentially by base pairing with U2 snRNA^{5,10,11}. In yeast there is an invariant branch point sequence, UACUAC, which base pairs with U2 snRNA (ref. 10). Increasing the base pairing potential between mammalian branch points and U2 snRNA improves their efficiency^{12–15}. Nevertheless, many mammalian branch points bear little

resemblance to the consensus; several lines of evidence, however, have led to the belief that the mammalian branch point is specified primarily by its proximity to the 3' splice site, possibly by a device that measures distance from the AG dinucleotide at the 3' intron boundary^{5,16,17}. First, most mammalian branch points have been mapped within 18–40 nucleotides of the 3' splice site^{5,16,18}. Second, mutation of branch points can lead to activation of cryptic sites within the same distance^{16,17}. Third, some introns lack A residues within 18–40 nucleotides of the 3' splice site, and use U or C residues instead, despite the clear preference for A residues¹⁹. In addition, it has been shown that, to variable extents, mutation of the 3' splice site AG can inhibit the first step in the splicing reaction, even though the 3' splice site is not itself cleaved until the second step^{18,20–23}.

Although the definition of the 5'-splice-site consensus sequence and its recognition by U1 RNA is well characterized^{8,24–26} the mechanism of recognition, relative importance and interactions of the 3' splice site elements, including the branch point, polypyrimidine tract and the 3' splice site AG, are poorly understood. The recent reports of branch points at distances of >140 nucleotides from the 3' splice site^{27–29}, clearly indicates that simple proximity to the 3' splice site cannot be an obligatory determinant of branch-point location.

We have now studied the mechanism of branch-point selection in the intron between the mutually exclusive exons 2 and 3 of the rat α -tropomyosin gene (*TM*). This intron has a branch point that lies 172 nucleotides upstream of the 3' splice site (Fig. 1a). The proximity of this branch point to the 5' splice site of exon 2 enforces mutually exclusive behaviour, because splicing factors are unable to bind productively to the two elements and form an active spliceosome^{27,28}. Using this *TM* intron and the first intron of the human β -globin gene, we show that the branch point is specified independently of the 3' splice site, by its immediate sequence context and by proximity to an adjacent polypyrimidine tract which may be distant from the 3' splice site AG. Efficient spliceosome formation, cleavage at the 5' splice site and lariat formation require the polypyrimidine tract and branch-point sequence, but can occur in the complete absence of the 3' splice site and exon. The 3' splice site itself is only recognized in the second step of splicing, and is in fact determined with reference to the branch point by a distance-independent scanning process for the first AG dinucleotide downstream of the branch point/polypyrimidine tract.

Spliceosome formation and 5' splice site cleavage

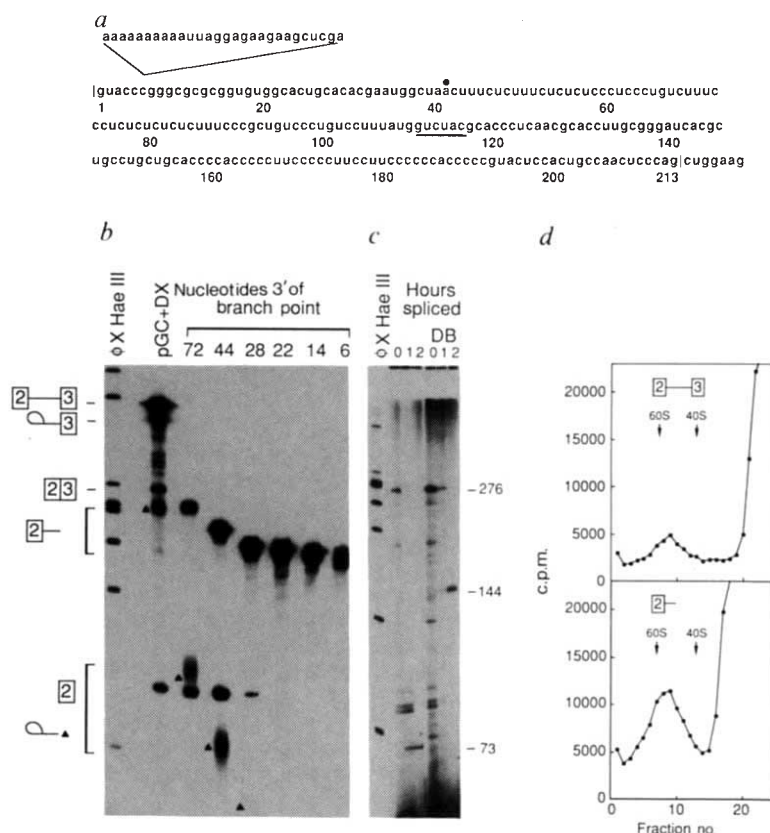
To investigate the factors determining the branch point of the *TM* intron between exons 2 and 3, transcripts were produced from construct pGC+DX (refs 27, 28), which contains a 28-nucleotide spacer adjacent to the 5' splice site (Fig. 1a). Transcripts lacking the final 100 nucleotides of the intron and the 3' exon were produced by linearizing the template at the *AccI* site in the middle of the intron (underlined in Fig. 1a). These transcripts underwent efficient 5' cleavage and lariat formation (Fig. 1b). The branch point used in lariat formation was in the same position as previously mapped, 41 nucleotides downstream of the 5' splice site (Fig. 1c; ref. 27). Moreover, analysis of the splicing reaction of the truncated transcript by glycerol gradient centrifugation revealed efficient formation of a 60S spliceosome peak (Fig. 1d) containing all of the principal snRNAs U1, U2, U4, U5 and U6 (data not shown). This indicates that interaction of U5 snRNP with the 3' splice site AG (ref. 30) is not essential for spliceosome formation. The high relative accumulation of the 60S complex formed by the truncated transcript compared with that formed by the full-length transcript presumably results from a block in the second step of splicing owing to the lack of a 3' splice site. Thus, spliceosome formation, correct branch-point selection and 5' splice site cleavage are independent of the 3' splice site. This seems to be a general principle of splicing, because similar findings have been reported for yeast introns³¹ and other mammalian introns (see below). The relatively passive role of the 3' splice site AG was further demonstrated in RNase H protection experiments⁴³, which showed stable interaction of nuclear factors with the 5' splice site, branch point and polypyrimidine tract, but not at the 3' splice site (data not shown). The progressive truncation of the *TM* transcript from the 3'

end, with a corresponding shortening of the polypyrimidine tract, led to a reduced efficiency of both cleavage at the 5' site and formation of the 60S complex; both were abolished when less than 28 nucleotides remained 3' of the branch point (Fig. 1b). This effect probably reflects the requirement for a minimal transcript size downstream of the branch point, because more extensive internal polypyrimidine deletions were required to abolish splicing activity (see below). The inhibition of 5' cleavage in some mammalian introns when the 3' exon has been deleted could also reflect an insufficient transcript size 3' of the branch point^{20,32}.

Associated branch point and polypyrimidine tract

Unlike the variable effects reported for 3' splice site AG mutations^{18,20-23}, deletions of 3' splice site-associated polypyrimidine tracts always inhibit the assembly of the splicing complex and the first step in splicing^{18,20,21}. By using *TM* transcripts with internal deletions, we demonstrated that the polypyrimidine tract adjacent to the branch point, and distant from the 3' splice site, is required for splicing activity (Fig. 2a) and complex formation (data not shown). Reduction of the pyrimidine tract from >50 to 17 nucleotides ($\Delta 66-110$) had no detectable effect on splicing or 60S complex formation. At 9 nucleotides ($\Delta 52-110$), both splicing and complex formation were significantly reduced, and with only one pyrimidine, both were abolished. Thus, the pyrimidine tract adjacent to the branch point is required for splicing, although it can be reduced in length before splicing is impaired. Similar incremental reductions in splicing efficiency of an intron of the rabbit β -globin gene accompanied progressive deletions of the polypyrimidine tract³³. In *TM* transcripts with the maximal pyrimidine-tract deletion ($\Delta 44-110$),

FIG. 1 5' splice site cleavage, branch-point selection and spliceosome formation occur in the absence of the 3' splice site and exon. a, Sequence of the α -*TM* intron used. Intron-exon boundaries are shown by vertical lines, and the sequence of the 28-nucleotide (nt) spacer element in the *SmaI* site at position 6 is shown. The branch point A residue is marked by a dot and the *AccI* site is underlined. b, *In vitro* splicing of transcripts with 3'-end truncations. The pGC+DX template used for transcription contained α -*TM* exons 2 and 3, the intact intron and the 28-nt spacer adjacent to the 5' splice site²⁷. Truncated transcripts were produced by templates terminated at the *AccI* site (72, underlined in a) or at *AccI* sites introduced closer to the branch point by oligonucleotide-directed mutagenesis (see Fig. 2, constructs p $\Delta 66-110$ to 44-110). The 3' termini of transcripts 44-6 are changed from the sequence shown in a only by the 3' GUCU of the introduced *AccI* sites. ³²P-labelled transcripts were made by *in vitro* transcription, spliced in HeLa cell nuclear extracts and fractionated by denaturing gel electrophoresis, as previously described²⁷. Standard splicing reactions were in 10 μ l with 10-30 fmol RNA, 6 μ l nuclear extract, 3.2 mM MgCl₂, 0.5 mM ATP, 20 mM creatine phosphate, 8 U RNasin. Two-hour time points are shown. c, Correct branch-point selection in the absence of the 3' splice site. Splicing reactions were performed with transcripts terminated 72 nt 3' of the branch point. Transcripts were spliced for the indicated times and primer extension of the RNA by reverse transcriptase was carried out directly, or after debranching (DB), using a 5' ³²P-labelled primer which hybridizes with positions 88-115 of the intron²⁷. Bands of 276, 73, and 144 nt represent extension to the 5' end of the unspliced transcript, to the branch point, and to the 5' end of the intron in debranched samples, respectively. d, Glycerol gradient centrifugation of splicing reactions. Full length or truncated transcripts (50 fmol), terminated at the *Bam*HI site 3' of exon 3 (top) or at the *AccI* site in the intron (bottom), respectively, were spliced *in vitro* for 1 h (25 μ l reactions) and then fractionated on 10-30% glycerol gradients at 40,000 r.p.m. for 3 h. The gradient contained 20 mM HEPES buffer pH 7.9, 1 mM MgCl₂, 60 mM KCl. Cerenkov counts of fractions



were measured. The positions of 60S and 40S fractions, determined by parallel centrifugation of HeLa cell ribosomes, are indicated.

an extensive stretch of pyrimidines was still present 50–60 nucleotides downstream of the branch point (positions 162–211; Fig. 1a). Further deletions bringing the branch-point sequence adjacent to this polypyrimidine tract restored splicing activity (Fig. 2a; $\Delta 44$ –157). This shows that a functional polypyrimidine tract does not have strict sequence specificity, and that to be active it requires an adjacent acceptable branch-point sequence.

Branch-point sequence context

Branch point formation can occur at sequences that are highly divergent from the consensus sequence^{16,17}. Nevertheless, A residues, such as those at positions 156, 161 and 186 in the *TM* intron (Fig. 1a), which have a suitable downstream pyrimidine tract, are not used as branch points under any conditions tested. To test the role of sequence context in preventing the use of such nucleotides as branch points, the sequence around the A residue 28 nucleotides upstream of the 3' splice site was converted to the strong branch-point consensus UACUAAC (ref. 13). This mutation resulted in splicing of exon 2 to exon 3 in transcripts in which the normal branch point is too close to the 5' splice site to be involved in branch formation (Fig. 2b; UACUAAC 181). Thus the pyrimidine environment downstream of position 186 is sufficient to support splicing, given a suitable adjacent branch-point sequence. The normal sequence at this position, CCCCCAC, is not an acceptable branch-point sequence, despite its having only two mismatches from the YNYURAY consensus. The replacement of the five purines between the UACUAAC and the 3' splice site with pyrimidines, thereby creating a 25-nucleotide continuous polypyrimidine tract, dramatically increased the efficiency of splicing and of 60S complex formation (UACUAAC 181Y). Thus, the sequence context of the branch point, and an adjacent polypyrimidine tract are involved in branch-point specification, and in determining the efficiency of the splicing reaction.

Scanning for the 3' splice site

Given that the specification of the branch point is independent of the 3' splice site AG, the question arises as to the specification of the 3' splice site itself. Inspection of the *TM* intron shows that the 3' splice site is the first AG dinucleotide downstream of the branch point and polypyrimidine tract (Fig. 1a), as it is in other introns^{5,29}. Thus, the 3' splice site could simply be recognized by 5'-to-3' scanning for an AG, as originally suggested^{6,34}. To test this possibility, several mutants were constructed in which AG dinucleotides were inserted between the branch point and the 3' splice site (Fig. 3). In all cases, the intervening AG was used exclusively as the 3' splice site. The nature of the nucleotide preceding the AG, however, had a profound effect on the efficiency of the second step in splicing; CAG, UAG and AAG trinucleotides were used efficiently as 3' splice sites (constructs ppR, p23Bg, p23Af). By contrast, GAG was inefficiently cleaved, although it was efficiently recognized, because downstream AG dinucleotides, including the normal 3' splice site, were not used. Moreover, a CAG-to-GAG mutation (pGAG; Fig. 3) severely reduced the rate, but not the position, of cleavage at the normal 3' splice site. These results are compatible with the absence of GAG trinucleotides at natural 3' splice sites^{1,5,7}. Deletion of the normal 3' splice site AG (p Δ AG) resulted in the exclusive and efficient use of the first AG in the 3' exon, 178 nucleotides downstream of the branch point. The distance between the branch point and the AG in these transcripts (75–178 nucleotides) had no effect on the efficiency of the second step of splicing (compare p23Bg and p Δ AG). All the AG dinucleotides used in the above transcripts had a pyrimidine-rich environment immediately upstream, which could have been involved in specification of the 3' splice site. This is unlikely because conversion of the 30 nucleotides immediately upstream of the 3' splice site in construct pGC+DX from 70% pyrimidine to 70% purine had no effect on either step in splicing (pR2, Fig. 3). This confirms that the polypyrimidine tract is not obligatorily

associated with the 3' splice site. Rather, it seems that the 3' splice site is recognized by a simple scanning process for the first AG downstream of the branch point. This is in agreement with the behaviour of some splicing mutants that use AG dinucleotides upstream of the normal 3' splice site^{18,34–37}. One possible caveat is that the AG may need to be beyond a minimal distance (>18 nucleotides) from the branch point^{5,34,38}, presumably because of the steric constraints of factors bound at the branch point and polypyrimidine tract. Also, in some cases more than one of a series of closely spaced AG dinucleotides is used as a 3' splice site (ref. 39; and R. Reed, personal communication). Nevertheless, in these cases it is the upstream site that is predominantly used³⁹.

In a preliminary investigation of the nature of the scanning process, we tested whether stable secondary structure between the polypyrimidine tract and the 3' splice site interferes with the recognition of the 3' splice site. Insertion of a 48-nucleotide palindromic element downstream of the polypyrimidine tract resulted in a block in the second step of splicing (p23HP, Fig. 3). The presence of stable secondary structure was confirmed by the resistance to digestion with RNase H in the presence of

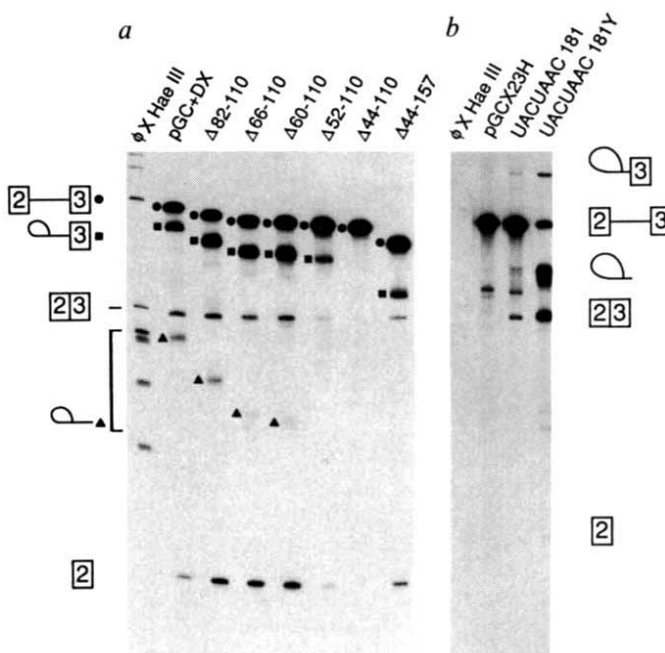


FIG. 2 Branch-point specification by adjacent polypyrimidine tract and by sequence context. *a*, Effect of polypyrimidine-tract deletions. Templates containing specific intron deletions were constructed by deleting between the *AccI* site at position 110 (see Fig. 1a) and *AccI* sites introduced by oligonucleotide directed mutagenesis with polymerase chain reaction (PCR) amplification, at the indicated positions. The structure of mutagenized and amplified DNA was verified by sequencing. The templates all contained the spacer element at the 5' end of the intron, as in plasmid pGC+DX. Splicing reactions were carried out and analysed as in Fig. 1b. Two-hour splicing reactions are shown. *b*, Branch-point sequence context is important for splicing. Constructs containing the native non-splicing intron with no spacer element at the 5' end (pGCX23H), and with the sequence context of the A residue at position 186 changed from CCCCCAC to UACUAAC (UACUAAC 181 and UACUAAC 181Y) were used as templates for transcription. UACUAAC 181Y also had all purines between the 3' splice site and UACUAAC sequence converted to pyrimidines (U at positions 192, 199 and 205; C at positions 194, 202 and 206). The products of 2-h splicing reactions are shown. The only band appearing in the plasmid pGCX23H lane corresponds to a block to endogenous exonuclease activity by factors bound at the branch point²⁷. The identities of lariat bands produced by the UACUAAC-element-containing transcripts were verified by debranching reactions.

Generality of specification mechanisms

To test the generality of our findings with the α -TM intron we made several constructs by using the first intron of the human β -globin gene. A 120-nucleotide spacer element derived from the α -TM intron, chosen because it lacks any cryptic branch points or 3' splice sites²⁷, was inserted between the globin gene polypyrimidine tract and 3' splice site (Fig. 4, BP BAM). This intron was still spliced, albeit with reduced efficiency. The correct globin splice sites and branch point, now 155 nucleotides upstream of the AG, were used, showing that the branch point need not be in proximity to the 3' splice site. The reduced splicing efficiency could be due to the increased distance of the branch

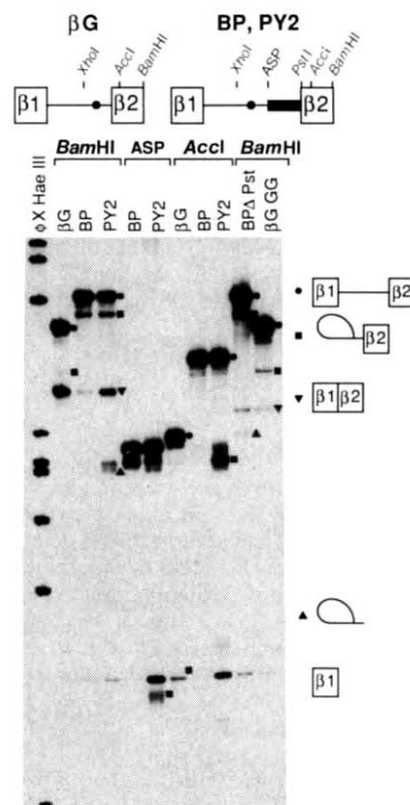


FIG. 4 Branch point and 3' splice site selection in β -globin transcripts. Transcripts were produced containing the first intron of the human β -globin gene and the flanking exons (β G). Template BP was derived from the parent β G-containing construct RR108 by inserting nucleotides 111-213 of the α -TM intron (shown by the thick line in the schematic diagram) between the β -globin polypyrimidine tract and 3' exon. Template BP was constructed by PCR amplification of a fragment from the *Xho*I site upstream of the β -globin branch point, up to, but not including the 3' splice site AG, followed by an ASP718 and an *EcoRV* site, nucleotides 111-213 of the TM intron with the final 6 nt changed to a *Pst*I site, and globin exon 2 up to the *Acc*I site. The *Xho*I-*Acc*I fragment was cloned back into the parent globin framework. Construct PY2 was derived from BP by the replacement of the six purines between the β -globin branch site and the ASP718 site with T residues, by oligonucleotide-directed mutagenesis with PCR amplification. Construct BP Δ Pst had the 3' splice site of BP removed by blunt-ending the *Pst*I site and re-ligating. Construct β G GG was obtained by deleting between the *Kpn*I (ASP) and *Pst*I sites of BP leaving a single A-to-G change in the 3' splice site of β G. Transcripts were produced from templates linearized at the *Bam*HI, *Acc*I and ASP718 sites. The products of 2-h splicing reactions are shown. The identities of all lariat bands were verified by debranching and the branch point used by β G and BP (shown by the dot) was verified by primer extension with reverse transcriptase.

point from the 3' splice site and the globin exon sequences, both of which have been shown to have a positive effect on splicing efficiency^{18,44}. Improvement of the globin polypyrimidine tract in the BP transcript by the replacement of all purines between the branch point and the spacer element to create a 34-nucleotide continuous pyrimidine stretch (PY2 Bam), resulted in highly efficient splicing. Truncation of the BP and PY2 transcripts between the polypyrimidine tract and 3' splice site resulted in cleavage at the 5' splice site (BP ASP, PY2 ASP), as with the *TM* transcript (Fig. 1a). The efficiency of 5' cleavage of BP, however, was severely inhibited, whereas PY2 was processed as efficiently as the full-length transcript. The efficiency of 5' cleavage was improved slightly in BP transcripts terminated at the *AccI* site, 14 nucleotides into exon 2, although it was still less efficient than the full-length transcript. Again, PY2 was efficiently processed, indicating that a strong polypyrimidine tract, as in the *TM* intron, may be able to ameliorate any negative effects otherwise brought about by the removal of exon sequences or the 3' splice site AG (refs 18, 44; and R. Reed, personal communication). In support of a scanning process in the recognition of the 3' splice site, deletion of 4 nucleotides in the BP transcript that includes the 3' splice site AG resulted in the use of the first AG in the globin exon 2, 183 nt downstream of the branch point, with comparable efficiency to the non-mutated BP. The same AG was used as a 3' splice site when an AG-to-GG mutation was made in the parent globin construct (β GG). In the latter case, however, the efficiency of the first step in splicing was reduced, consistent with a positive influence of a branch-point proximal AG on splicing efficiency. Previous reports have stressed the overall reduced splicing efficiency and the lack of cleavage at mutated 3' splice sites^{18,20-23}, rather than the use of the next available AG. Nevertheless, a mutant rabbit β -globin which did not splice *in vitro* was shown to use the first AG in the exon *in vivo*²². Thus the β -globin data are consistent with those from the *TM* transcripts. These results also confirm, however, that in the absence of a strong polypyrimidine tract, a nearby 3' splice site AG can improve the efficiency of early steps in the splicing reaction. Candidate factors for AG recognition include an intron binding protein of relative molecular mass (M_r) 70,000–100,000^{47,48} and a heterogeneous nuclear ribonucleoprotein (hnRNP) A1 protein⁴⁹. When the separation of the branch point and the 3' splice site is not large, interaction with the same factor(s) may occur before downstream scanning, stabilizing other transcript-spliceosome interactions, and accounting for the positive effect of AG dinucleotides close to the branch point in some transcripts (Fig. 4; refs 18 20–23).

Discussion

Our results clarify the relationship between the splicing elements at the 3' end of mammalian introns. Spliceosome assembly and the first step in splicing require the 5' splice site, the branch point and the polypyrimidine tract. The 3' splice site AG is not required before the second step in splicing, when it is located by scanning from the branch point/polypyrimidine tract.

The potential for the control of alternative splice-site selection by the branch-point sequence and polypyrimidine tract is realized in the case of *TM* exons 2 and 3. The 'default' selection of exon 3 in most cells is due to both the extensive polypyrimidine tract and to the high match of the branch-point sequence to the consensus. These easily out-compete the equivalent exon-2-associated elements (M. Mullen and B. N-G., unpublished observations). Whereas consensus branch points are known to favour base pairing with U2 snRNA¹¹⁻¹⁵, the extensive exon 3 polypyrimidine tract seems to provide a strong binding site for the polypeptide of M_r 65,000 of U2AF (J. G. P., unpublished observations; and P. Zamore and M. R. Green, personal communication), a factor required for binding of U2 snRNP to the branch point⁴⁵. Likewise, the cell-specific switching of 3'-splice-site selection probably involves regulation targeted at the branch point and polypyrimidine tract rather than at the 3' splice site itself. The minimal sequence requirements for functional branch points and polypyrimidine tracts are at present not clear. Nevertheless, the use of a repetitive GU element as the functional substitute of the polypyrimidine tract in at least one intron⁴⁶ points to considerable flexibility in these requirements.

Scanning for the 3' splice site seems to have no maximal distance constraints other than the necessity to avoid the disruption of open reading frames by the use of intervening AG dinucleotides as cryptic splice sites. Very large separation of branch points and 3' splice sites may only occur when extensive polypyrimidine tracts are present^{27,29}, because by their nature these preclude the presence of intervening AG dinucleotides. The specification of the 3' splice site only as an AG downstream of the branch point could be an important factor in the evolution of exon sequences. The relative ease of creating new 3' splice site AG dinucleotides suggests that the incorporation of new coding segments is more probable at the 5' end of exons. This process has already been shown to underlie certain genetic disorders in which protein function is disrupted either by in-frame insertion of peptide segments³⁶ or by frameshifting of the normal exon sequence^{35,37}. Such mutations could sometimes be selectively advantageous, and thus provide an important source of protein sequence diversity. □

Received 3 August; accepted 5 October 1989

- Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. & Sharp, P. A. *Rev. Biochem.* **55**, 1119–1150 (1986).
- Brody, E. & Abelson, J. *Science* **228**, 963–967 (1985).
- Grabowski, P. J., Seller, S. R. & Sharp, P. A. *Cell* **42**, 345–353 (1985).
- Maniatis, T. & Read, R. *Nature* **325**, 673–678 (1987).
- Green, M. R. *Rev. Genet.* **20**, 671–708 (1986).
- Mount, S. M. *Nucleic Acids Res.* **10**, 459–472 (1982).
- Onshima, Y. & Gotch, Y. *J. molec. Biol.* **195**, 247–259 (1987).
- Zhuang, Y. & Weiner, A. M. *Cell* **46**, 827–835 (1986).
- Black, D. L., Chabot, B. & Steitz, J. A. *Cell* **42**, 737–750 (1985).
- Parker, R. & Guthrie, C. *Cell* **41**, 107–118 (1985).
- Keller, E. B. & Noon, W. A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7417–7420 (1984).
- Reed, R. & Maniatis, T. *Genes Dev.* **2**, 1268–1276 (1988).
- Zhuang, Y., Goldstein, A. M. & Weiner, A. M. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2752–2756 (1989).
- Zhuang, Y. & Weiner, A. M. *Genes Dev.* **3**, 1545–1552 (1989).
- Wu, J. & Manley, J. L. *Genes Dev.* **3**, 1553–1561 (1989).
- Ruskin, B., Greene, J. M. & Green, M. R. *Cell* **41**, 833–844 (1985).
- Padgett, R. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 8349–8353 (1985).
- Reed, R. & Maniatis, T. *Cell* **41**, 95–105 (1985).
- Hartmuth, K. & Barta, A. *Molec. cell Biol.* **8**, 2011–2020 (1988).
- Ruskin, B. & Green, M. R. *Nature* **317**, 732–734 (1985).
- Freudenberg, D. & Keller, W. *Cell* **42**, 355–367 (1985).
- Aebi, M., Hornig, H., Padgett, R. A., Reiser, J. & Weissman, C. *Cell* **47**, 555–565 (1985).
- Lamond, A. I., Konarska, M. M. & Sharp, P. A. *Genes Dev.* **1**, 532–543 (1987).
- Nelson, K. K. & Green, M. R. *Genes Dev.* **2**, 319–329 (1988).
- Mayeda, A. & Onshima, Y. *Molec. cell Biol.* **8**, 4484–4491 (1988).
- Eperon, L. P., Estibeiro, J. P. & Eperon, I. C. *Nature* **324**, 280–282 (1986).

- Smith, C. W. J. & Nadal-Ginard, B. *Cell* **56**, 749–758 (1989).
- Smith, C. W. J. *et al. in Cellular and Molecular Biology of Muscle Development* (eds, Keddes, L. K. & Stockdale, F. E.) 573–584 (Liss, New York, 1989).
- Helfman, D. M. & Ricci, W. M. *Nucleic Acids Res.* **17**, 5633–5650 (1989).
- Chabot, B., Black, D. L., LeMaster, D. M. & Steitz, J. A. *Science* **230**, 1344–1349 (1985).
- Rymond, B. C. & Rosbash, M. *Nature* **317**, 735–737 (1985).
- Furdon, P. J. & Kole, R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 927–931 (1986).
- Wieringa, B., Hofer, E. & Weissmann, C. *Cell* **37**, 915–925 (1984).
- Langford, C. J., Klinz, F. J., Donath, C. & Gallwitz, D. *Cell* **36**, 645–653 (1984).
- Spritz, R. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 2455–2459 (1981).
- Nakano, T. & Suzuki, K. *J. biol. Chem.* **264**, 5155–5158 (1989).
- Metherall, J. E., Collins, F. S., Pan, J., Weissman, S. M. & Forget, G. *EMBO J.* **5**, 2551–2557 (1986).
- Fu, X.-Y., Colgan, J. D. & Manley, J. L. *Molec. cell Biol.* **8**, 3582–3590 (1988).
- Fu, X.-Y., Ge, H. & Manley, J. L. *EMBO J.* **7**, 809–817 (1988).
- Konarska, M. M., Padgett, R. A. & Sharp, P. A. *Cell* **42**, 165–171 (1985).
- Halford, H. & Gallwitz, D. *Nucleic Acids Res.* **16**, 10413–10423 (1988).
- Cellini, A., Felder, E. & Rossi, J. J. *EMBO J.* **5**, 1023–1030 (1986).
- Ruskin, B. & Green, M. R. *Cell* **43**, 131–142 (1985).
- Reed, R. & Maniatis, T. *Cell* **46**, 681–690 (1986).
- Ruskin, B., Zamore, P. D. & Green, M. R. *Cell* **52**, 207–219 (1988).
- Shelley, C. S. & Baralle, F. E. *Nucleic Acids Res.* **15**, 3787–3799 (1987).
- Tazi, J. *et al. Cell* **47**, 755–766 (1986).
- Gerke, V. & Steitz, J. A. *Cell* **47**, 973–984 (1986).
- Swanson, M. S. & Dreyfuss, G. *EMBO J.* **7**, 3519–3529 (1988).

ACKNOWLEDGMENTS. We thank Dola Sengupta and Roger Breitbart for critical reading of the manuscript, Emily Flynn-McIntosh for artwork, Shahla Movahedi for oligonucleotide synthesis and Robin Reed for the plasmid RR108. This work was supported in part by the NIH and the Muscular Dystrophy Association of America.

Crystal structure of chaperone protein PapD reveals an immunoglobulin fold

Anders Holmgren & Carl-Ivar Brändén

Department of Molecular Biology, Biomedical Centre Box 590, S-751 24 Uppsala, Sweden

The chaperone protein PapD mediates assembly of pili in *Escherichia coli*. Its polypeptide chain folds into two immunoglobulin-type domains that are homologous in sequence to the human lymphocyte differentiation antigen Leu-1/CD5.

UROPATHOGENIC *Escherichia coli* carries an operon of 11 genes¹⁻³ for synthesis and assembly of pap pili which mediate attachment of the bacteria to eucaryotic cell surfaces. One of these genes, *PapD*, codes for a chaperone protein which mediates assembly of pili by forming soluble multimeric complexes with pili subunits as an intermediate step in the assembly process⁴. We report here the crystal structure of PapD to 2.5 Å resolution.

The crystal structure of PapD

The PapD protein was crystallized in an orthorhombic space group $P2_12_12_1$ with cell dimensions $a = 58.2$ Å, $b = 64.0$ Å and $c = 67.0$ Å, with one molecule in the asymmetric unit⁵. The processed protein was purified from an *E. coli* expression system and included the region of the gene encoding the signal peptide. X-ray data to 2.5 Å resolution were collected on a Xenotronics area detector. A multiple isomorphous replacement (MIR)/anomalous electron density map was calculated using three heavy metal derivatives giving an average figure of merit of 0.66. Phasing statistics are given in Table 1. The polypeptide chain was traced using the Bones option in Frodo⁶. The amino-acid sequence⁴, as predicted from the complementary DNA sequence, was fitted into the map using the aromatic side chains as reference points (Fig. 1). The structure has been initially

refined using Xplor⁷ on an Alliant Fx/4 computer giving an *R*-factor of 0.24 with no water molecules added. The current model of 218 residues leaves no regions of the MIR map unassigned. The amino acids of the model are numbered from 22 to 239 to conform to the numbering of the published sequence with the leader peptide (1-21) included. Coordinates of the current model have been deposited in the Brookhaven Protein Data Bank.

The structure of PapD is built up from two globular domains oriented such that the complete molecule has the shape of a boomerang (Figs 2 and 4). The two domains have very similar structures with an antiparallel beta-barrel structure formed by two beta sheets packing against each other. The topology of each domain is the same as the constant domain of immunoglobulin molecules, comprising seven anti-parallel beta strands, A to G (ref. 8). Both domains lack the disulphide bridge between beta strands B and F which is usually but not always found in immunoglobulin domains⁸. The residues that form the beta strands are given in Fig. 2b.

The N-terminal domain comprises residues 22-121. Two long loop regions are present at one end of the barrel between strands C and D (residues 62 to 73) and between strands E and F (residues 94-106). The other loop regions are either reverse turns or short loops comprising three to five residues.

Residues 153 to 239 form the C-terminal domain. In addition to the seven beta strands A to G of the immunoglobulin fold there is a short beta strand H at the C-terminal end of the polypeptide chain which is positioned between, and is anti-parallel to, beta strands G and A (Fig. 2). This beta strand is firmly linked to strand G by a disulphide bridge between Cys 228 and Cys 233. The four residues between these cysteines

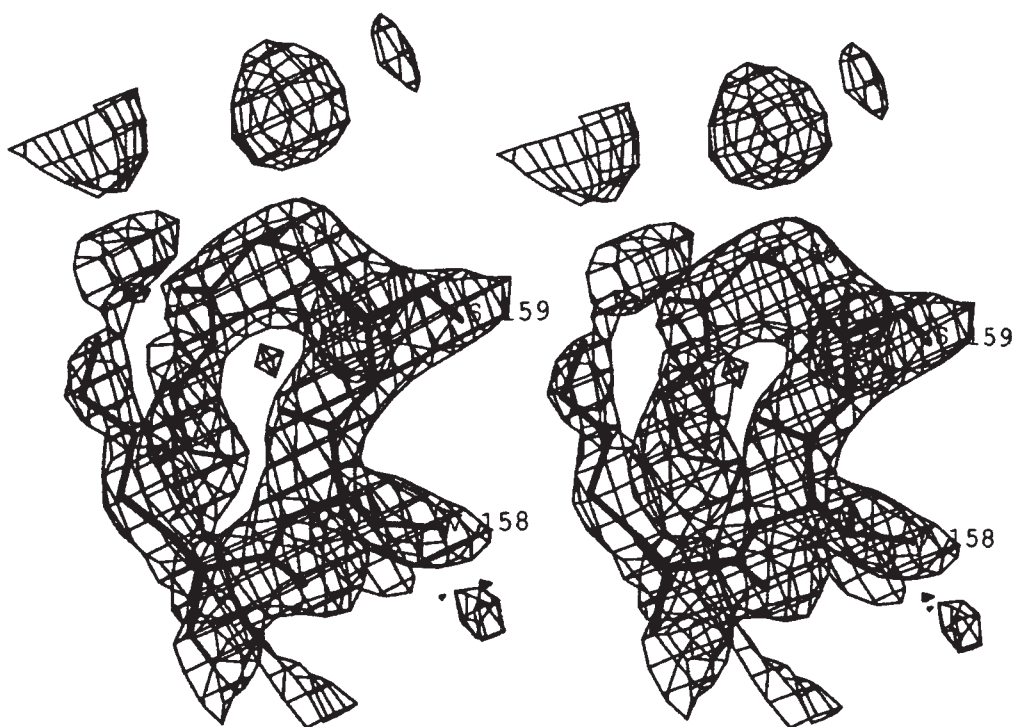


FIG. 1 Representative electron density map showing a reverse beta-turn between strands A and B in the C-terminal domain.

TABLE 1 Crystallographic data

Space group P 2 ₁ 2 ₁ 2 ₁ (<i>a</i> =58.2 Å, <i>b</i> =64.0 Å, <i>c</i> =67.0 Å)									
Data set	Native		K ₂ PtCl ₄		K ₂ Pt(SCN) ₄		Pt(NH ₃) ₂ Cl ₂		
No. of observations	20,820		20,429		20,375		15,100		
No. of unique reflexions	6,353		6,456		7,630		6,633		
Possible reflexions (%)	69.2		69.2		81.4		68.4		
<i>R</i> merge (%)	5.3		5.5		6.6		7.4		
Mean fractional isomorphous change (%)	—		20.4		15.5		22.3		
No. of heavy atom sites	5		4		5		5		
<i>F_h</i> / <i>E</i> r.m.s.	3.20		1.26		2.19		2.19		
<i>R</i> cullis	0.46		0.78		0.51		0.51		
Resolution (Å)	17.0	9.3	6.4	4.9	3.9	3.3	2.8	2.5	total
No. of reflections	35	169	387	682	1,011	1,363	1,587	819	6,053
Figure of merit	0.81	0.89	0.89	0.82	0.73	0.66	0.56	0.45	0.66

$R \text{ merge} = \sum_i (I_i - \bar{I}) / \sum_i \bar{I}$
 F_h/E r.m.s. F_h =heavy atom structure factor, E is the residual lack of closure
 $R \text{ cullis} = \sum |F_{ph}(\text{obs}) - F_{ph}(\text{calc})| / \sum |F_{ph}(\text{obs}) - F_p(\text{obs})|$

PapD was isolated from the periplasm of an *E. coli* strain containing a cloned *papD* gene under the influence of the *tac* promoter⁴, thereby forcing the bacteria to overproduce the protein. PapD was collected from the periplasmic space and purified by the method of Lindberg *et al.*⁴ with some modifications⁵. Crystals of PapD were obtained by the hanging drop method of vapour diffusion using a protein concentration of 12 mg ml⁻¹ (ref. 5), and 15% polyethylene glycol 8000 as precipitating agent. Crystals appeared after a few days and usually grew to a size of 0.5 mm in all directions. X-ray data were initially collected to 5.0 Å on a single counter four circle diffractometer. Analysis of difference Patterson maps was undertaken using the program RSPS (written by Stefan Knight), which provided the heavy atom positions for the major site in the three derivatives used and indicated some minor sites of low occupancy. The heavy-atom positions were further refined using the program Heavy in the CCP4 program package from Daresbury UK. Difference Fourier maps, calculated by the program Protein²¹, revealed some additional minor sites. X-ray data to 2.5 Å resolution were collected for those derivatives which gave good refinement statistics on a Xentronics multiwire area-detector²², using the software system described by Blum *et al.*²³.

form a short hairpin loop between strands G and H. The sulphur atoms of this disulphide bridge (Fig. 3) are oriented towards the hydrophobic core of the barrel and do not react with heavy metal reagents. The linker region between the domains, residues 122 to 152 contains a number of hydrophobic side chains which form a hydrophobic core in the central region of the molecule. Additional residues in this hydrophobic core are provided by loop regions from both domains. Secondary structure elements in relation to the amino-acid sequence of PapD are given in Fig. 5.

Structural similarity with immunoglobulins

To determine whether the topological similarity between the PapD and immunoglobulin domains also reflects structural similarity, we have compared these domain structures using the HOMO program⁹. The alpha carbon atoms of the two PapD domains were superimposed with each other and with the constant domain CH2 of an IgG Fc-fragment¹⁰. We found that 50 alpha carbon atoms of the PapD domains superimpose with each other with a root mean square error (r.m.s.) of 3.5 Å, showing that only about half of the residues that build up the two domains are structurally equivalent. These residues build up the core of the domains. An r.m.s. value of 3.5 Å for superposition of the cores of two structurally equivalent domains is compatible with the fact⁴ that there is no amino-acid homology between these two domains¹¹.

In contrast to the above results, the cores of 50 alpha carbon atoms of PapD domains 1 and 2 superimpose with CH2 with r.m.s. values of 2.1 and 2.5 Å respectively. Each PapD domain is thus more similar in structure to the immunoglobulin domains than they are to each other. They seem to be two different variations of the CH2 core structure. In both these PapD domains the beta strands are arranged in two distinct beta sheets, similar to the arrangement in all immunoglobulin structures⁸ and different from that in other structures of the same topology such as superoxide dismutase, which is more like a barrel¹².

Sequence homology with CD5-antigen

There is no significant amino-acid sequence similarity between CH2 and the PapD domains. To examine if other members of

the immunoglobulin superfamily are homologous to PapD we searched in the NBRF protein-sequence data base (Protein Identification Resource), and found that amino-acid residues -4 to 252 of the Leu-1 human lymphocyte differentiation antigen also called CD5 (ref. 13), could be aligned with the sequence of the complete PapD molecule giving a 26% identity. All gaps

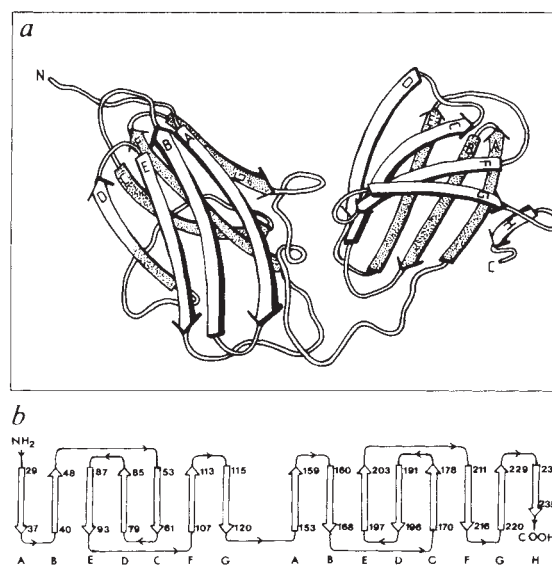


FIG. 2 *a*, Schematic diagram of the PapD molecule, showing the arrangement of the beta strands in the two domains. Strands A B E form one beta sheet and strands C F G form a second beta sheet which is packed against the first sheet. Dotted strands belong to the sheet at the back of the figure. Beta strand D of the N-terminal domain belongs to both beta sheets by forming hydrogen bonds to strand C at the beginning and to strand E at its end. *b*, Topology diagram of the course of the polypeptide chain of PapD. Residue numbers that form approximate beginnings and ends of the beta strands are given starting with residue 22 which is the first residue after the signal peptide.

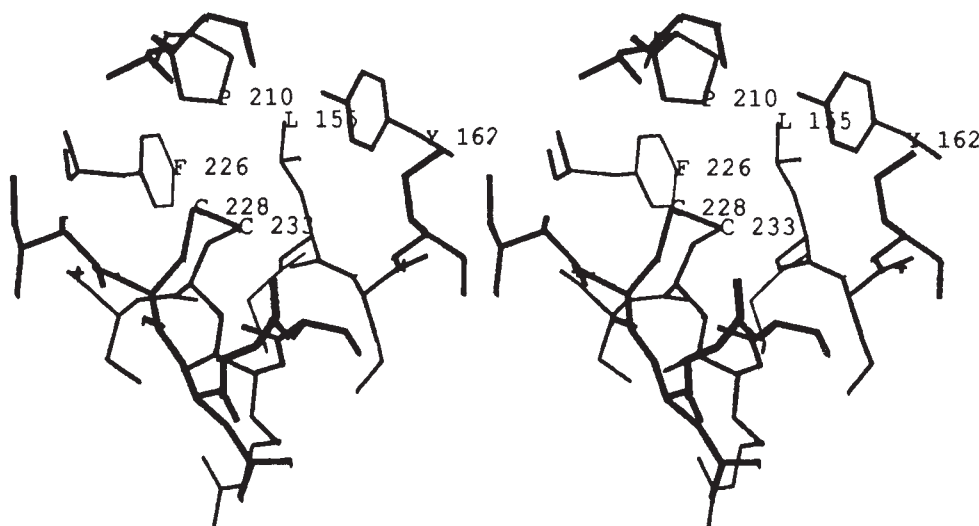


FIG. 3 Stereodigram of the only disulphide bridge in PapD (residues 228–233) and some surrounding residues.

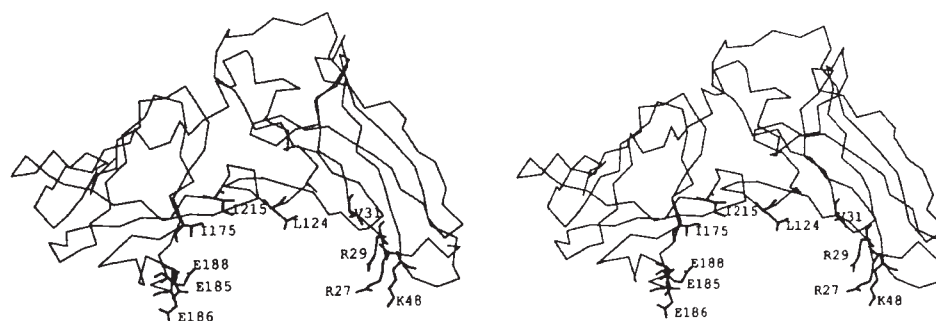


FIG. 4 Stereo image of the Ca-chain including side chains that might be important in binding other pilin subunits. The probable binding surface consists of an hydrophobic area at the cleft between the domains comprising residues (V31, L124, I175, I215). At each side of this surface there is a cluster of charged residues. The N-terminal part has a patch of three basic residues (R27, R29, K48) and the C-terminal domain three acidic residues (E185, E186, E188).

in this alignment occur either in loop regions between the beta strands or in the linker region between the PapD domains.

Domain 1 of CD5 (residues 1–90) has been claimed to have an immunoglobulin fold¹⁴ mainly based on the distribution of Cys-residues but this claim was queried¹⁵ as no significant sequence homology was found with members of the different subclasses V, C1 and C2 of the immunoglobulin superfamily. We have shown here that both PapD and CD5 belong to the immunoglobulin superfamily and we propose that they form a fourth subclass, C3. All three domains of Leu-1/CD5 (ref. 13) belong to this subclass, domain 1 (residues 1–109) and domain 2 (residues 132–247) because they are both homologous in sequence to PapD, and domain 3 (residues 250–320) because it is homologous to domain 1 of CD5.

Conclusions

PapD is a chaperone protein that mediates assembly of the pap pili fibre protein complex on the surface of *E. coli*. These pili are present on pathogenic *E. coli* isolates that are clinically associated with the disease pyelonephritis¹. The function of pap pili, to attach bacteria to epithelial cells in the urinary tract, is correlated with a minor subunit at the tip of the pili, PapG, which contains a specific binding site for carbohydrate receptors, gal- α (1–4)gal on the surface of human epithelial cells². PapD which is a periplasmic protein product of the *pap* operon, ensures the stability of the pili subunits by protecting them from aggregation and enzymatic degradation before they reach the assembly site at the outer membrane⁴. PapD is not part of the final pilus complex. Pre-assembly complexes of PapD with PapG as well as with another minor subunit PapE have been identified and purified³. Because PapD[–] mutants cannot express pili⁴ it is

apparent that PapD is required for correct assembly by forming pre-assembly complexes with pili subunits. A subclass of similar assembly proteins, GroE¹⁶, Hsp 60 (ref. 17) and Rubisco binding protein¹⁸ have been identified in other systems and are called chaperonins¹⁹. We do not yet know whether PapD also assists in the folding mechanism as Hsp 60 seems to do²⁰.

We have examined the surface of the PapD molecule to see if a reasonable guess can be made of the interaction area with pili subunits, for the preassembly complex. One region in the wide crevice between the domains has features which are not characteristic of protein surfaces (Fig. 4). The central part of this region contains a solvent-exposed hydrophobic patch comprising residues V31, L124, I175 and I215. On each side of this hydrophobic area there are regions with charged residues, one

```

*****A***** *****B*****
1  MIRKKILMAA IPLEVISGAD AAVSLDRTRA VFDGSEKSM TLDISNDNKLQ
*****C***** *****D***** *****E*****
51 PYLAQAWTEN ENQEKIITGP VIATPPVQRL DPGAKSMVRL STTPDISKLP
*****F***** *****G*****
101 QDRESLFYFN LREIPPRSEK ANVLQIALQT KIKLFYRPA A IKTRPNEVWQ
*****A***** *****B***** *****C***** *****D***** *****E*****
151 DQLILNKVSG GYRIENPTPY YVTVIGLGG S EKQAEEGEFE TVMLSPRSEQ
*****F***** *****G***** *****H*****
201 TVKSANYNTP YLSYINDYGG RVLFSFICNG SRCVSKKEK

```

FIG. 5 Amino-acid sequence of PapD as predicted from the cDNA sequence⁴. Residues 1–21 form the signal peptide which is cleaved off in the cells before purification of the protein. The model of the X-ray structure comprises residues 22–239. The top line identifies the beta strands, A to G in domain 1, and A to H in domain 2.

patch of three basic residues (R27, R29, K48) on the N-terminal domain and another area of three acidic residues on the C-terminal domain (E185, E186, E188). We suggest that this region is involved in binding the pili subunits. This model of the interaction area forms the basis for specific mutagenesis work, now in progress.

The amino-acid-sequence homology between *E. coli* PapD and human Leu-1/CD5 raises some puzzling evolutionary questions. The rate of evolution in the CD5 proteins is quite fast, as the functionally homologous CD5 proteins in human (Leu-1)

and mouse (Ly-1) show rather low sequence identity¹⁵, from 43% in domain 1 to 58% in domain 3. The possibility of divergent evolution from a common ancient bacterial ancestor therefore seems to be excluded. It is more likely that horizontal gene transfer has occurred at a much later stage in evolution. Genes involved in eukaryotic cell-cell interactions may have been recruited by bacteria to aid in their attachment to eukaryotic cells. Such gene transfer would also explain why a member of the immunoglobulin superfamily has been found in *E. coli* bacteria. □

Received 28 July; accepted 4 October 1989.

1. Norgren, M. *et al.* *EMBO J.* **3**, 1159–1165 (1984).
2. Lindberg, F. P. *et al.* *Nature* **328**, 84–87 (1987).
3. Hultgren, S. J. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **86**, 4357–4361 (1989).
4. Lindberg, F. P. thesis, Umeå Univ. (1987).
5. Holmgren, A., Bränden, C.-I., Lindberg, F. & Tennent, J. M. *J. molec. Biol.* **203**, 279–280 (1988).
6. Jones, T. A. & Thirup, S. *EMBO J.* **5**, 819–822 (1986).
7. Brunger, A. T., Kuriyan, J. & Karplus, M. *Science* **235**, 458–460 (1987).
8. Williams, A. F. & Barclay, A. N. *Rev. Immun.* **6**, 381–405 (1988).
9. Rao, S. T. & Rossmann, M. G. *J. molec. Biol.* **76**, 241–256 (1973).
10. Deisenhofer, J. *Biochemistry* **20**, 2361–2370 (1981).
11. Chothia, C. & Lesk, A. M. *EMBO J.* **5**, 823 (1986).
12. Richardson, J. S. *et al.* *J. molec. Biol.* **102**, 221–235 (1976).

13. Jones, N. H. *et al.* *Nature* **323**, 346–349 (1986).
14. Huang, H.-J. S. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 204–208 (1987).
15. Williams, A. F. *Immun. Today* **8**, 298–303 (1987).
16. Hemmingsen, S. *et al.* *Nature* **333**, 330–334 (1988).
17. McMullin, T. W. & Hallberg, R. L. *Molec. cell. Biol.* **8**, 371–380 (1988).
18. Goloubinoff, P. *et al.* *Nature* **337**, 44–47 (1989).
19. Ellis, R. J. *Nature* **328**, 378–379 (1987).
20. Ostermann, J. *et al.* *Nature* **341**, 125–130 (1989).
21. Steigemann, W. thesis, Technische Univ. Munich (1974).
22. Durbin, R. M. *et al.* *Science* **232**, 1127–1132 (1986).
23. Blum, M. *et al.* *J. appl. Crystallogr.* **20**, 235–242 (1987).

ACKNOWLEDGEMENTS. We thank Fredrik Lindberg and Staffan Normark for providing us with the *E. coli* strain containing PapD.

LETTERS TO NATURE

The distribution of galaxies in the direction of the 'Great Attractor'

Somak Raychaudhury

Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

THE discovery of galaxy streaming motions of up to 600 km s^{-1} towards Hydra–Centaurus has led to the suggestion¹ that this motion is induced by a concentration of galaxies (the 'Great Attractor') centred on galactic coordinates $l = 307^\circ$, $b = 9^\circ$ at a distance corresponding to $4,350 \pm 350 \text{ km s}^{-1}$ in the Hubble Flow. Here we present a map of $\sim 17,000$ galaxies down to $B_J = 17^m$ in an extended region of the sky in this direction. The clusters of galaxies identified from the map are shown to constitute two distinct concentrations in redshift (see, for example, ref. 2). We explore the possibility that the further of the two concentrations, centred near $l = 312^\circ$, $b = 31^\circ$, at a mean redshift of $14,000 \text{ km s}^{-1}$, is a major contributor to the peculiar motion of the Local Group. This is unlikely because the contribution to the dipole anisotropy of extragalactic light due to galaxies in this region, beyond a diameter limit of 1.3 arcmin , is about 10% of that due to nearer galaxies, and the concentration would need to have an extremely high mass-to-light ratio. Superclusters as massive as this would cause appreciable anisotropies in the cosmic microwave background radiation.

We have completed a survey of galaxies in an extended region around the proposed 'Great Attractor', using glass copies of UK Schmidt Telescope (UKST) Sky Survey plates (emulsion IIIa-J), which have been scanned by the SERC Automatic Photographic Measuring (APM) facility in Cambridge. The area covered (Fig. 1) consists of 103 such plates, each covering $5^\circ \times 5^\circ$ of the sky. After preliminary star-galaxy separation, all galaxy candidates have been visually inspected, and morphological types assigned to the galaxies found. The magnitudes of galaxies on different plates have been reduced to a consistent system by using the galaxies in the $\geq 1^\circ$ overlaps between neighbouring plates. A catalogue of galaxies (the Hydra–Centaurus catalogue), which is $\geq 97\%$ complete down to a magnitude limit of $B_J = 17^m$, has been compiled. It consists of 16,963 galaxies, for which positions, relative magnitudes, diameters and various profile para-

meters have been measured: $>80\%$ of these being catalogued for the first time.

The galaxies in the Hydra–Centaurus catalogue are plotted in an equal-area projection in Fig. 1a. All the Abell–Corwin–Olowin (ACO) clusters³ found in this region, along with some poor clusters, are indicated in Fig. 1b, and listed in Table 1. These figures can be compared with Fig. 2, which is a plot of all galaxies $>1 \text{ arcmin}$ in this region of the extragalactic sky, from existing catalogues (ESO, MCG and UGC: see Fig. 8 of ref. 1).

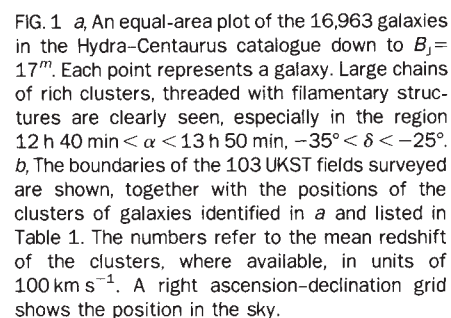
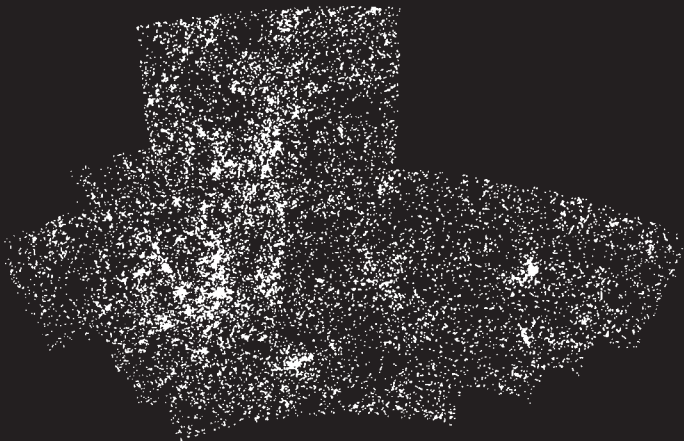
As large-scale redshift surveys in this area^{4,5} have concentrated mostly on galaxies in the ESO catalogue (ref. 15) not many redshifts are found in the literature for galaxies beyond $\sim 6,000 \text{ km s}^{-1}$. However, many clusters in this region out to $20,000 \text{ km s}^{-1}$ have been studied in detail^{6–10}. Figure 3 shows that almost all of these clusters belong to two distinct groups, with none between $5,000 \text{ km s}^{-1}$ and $10,000 \text{ km s}^{-1}$ in the area of interest. The nearer group is what is commonly known¹¹ as the 'Hydra–Centaurus Supercluster'. On the other hand, all the 14 clusters lying between $10,000 < v < 16,500 \text{ km s}^{-1}$ lie within 10° of A 3558 (Shapley 8: $l = 312^\circ$, $b = 31^\circ$), the mean redshift being $\langle v \rangle = 13,620 \text{ km s}^{-1}$. The existence of an excessive concentration of galaxies in this region was first discovered by Shapley¹², and is referred to here as the 'Shapley Concentration'. This concentration has recently been pointed out in the distribution of galaxies⁶, ACO rich clusters² and X-ray clusters¹³.

Because both gravity and luminosity fall as the inverse square of the distance, the dipole moment of the distribution of galaxies (weighted by a function of their luminosities) has often been used^{11,14} to estimate the peculiar acceleration of the Local Group. Using the expression $P = \sum_i D_i^2 \hat{r}_i$ as such an estimator, where D_i is the major angular diameter of a galaxy i , normalized to the ESO system¹⁵ and corrected for galactic extinction¹⁶, the 'optical dipole' has been estimated to be $P_{>1.3'} \approx 17,300$ (ESO arcmin)², directed towards $l = 257^\circ$, $b = 34^\circ$, from an all-sky catalogue of 16,886 galaxies having $D_{\text{ESO}} > 1.3'$ (ref. 14, Table 1).

Here, we attempt to estimate the contribution of the galaxies in the Hydra–Centaurus catalogue, beyond the above diameter limit, to the optical dipole. To do this, we have calculated the value of

$$\mathcal{P}_{<1.3'} = \sum_{\text{fields}} \left[\left(\sum_j D_j^2 \hat{r}_j \right) - \langle \Delta_{\text{SGP}}^2 \rangle \hat{r}_f \right],$$

For the 12,325 galaxies in the Hydra-Centaurus catalogue (103 UKST fields) satisfying $0.5' \leq D_{\text{ESO}} \leq 1.3'$, we derive $\mathcal{P}_{<1.3'} \approx 3,242$ (ESO arcmin)², directed towards $l = 310^\circ$, $b = 39^\circ$, which is 19% of $\mathcal{P}_{>1.3'}$. One can, however, argue that a significant


$$\mathcal{L}(d) = \frac{1+t/2}{(1+t/5)^4},$$

It has been shown that more than half of the dipole that is due to galaxies larger than $1'$ (ref. 11) comes from galaxies with $v < 2,000 \text{ km s}^{-1}$, and 80% from those with $v < 4,000 \text{ km s}^{-1}$. For the sake of estimating $\mathcal{L}^{-1}$, we assume that half of $\mathcal{P}_{>1.3'}$ comes from $\langle v \rangle = 2,000 \text{ km s}^{-1}$ and the other half from $\langle v \rangle = 3,500 \text{ km s}^{-1}$, which gives $\mathcal{L}^{-1} = 1.11$. This indicates that more than half of $\mathcal{P}_{<1.3'}$ as estimated above comes from smaller galaxies that are in the nearer concentration. Accordingly, a similar correction should be applied to the further concentration

as well, to account for galaxies missed by the diameter limit of the Hydra-Centaurus catalogue. Because the latter is magnitude limited, its diameter limit is not well defined; however, we believe that it is fairly complete down to $D_{\text{ESO}} = 0.5'$. Assuming that all the excess galaxies in the diameter range $0.5' \leq D_{\text{ESO}} \leq 1.3'$ that belong to the further concentration are at $\langle v \rangle = 13,620 \text{ km s}^{-1}$, we get $\mathcal{L}^{-1} = 1.42$. This means that the galaxies that are beyond a redshift of about $5,000 \text{ km s}^{-1}$ contribute only $1.42 \times (0.19 - 0.11)/1.11 = 10\%$ of the dipole acceleration that is due to those that are nearer.

The fact that the Shapley Concentration lies very close to the proposed position of the 'Great Attractor' has prompted some

authors^{2,6} to consider the possibility of its being a major contributor to the peculiar motion of the Local Group. Here, we consider the extreme hypothesis that the Shapley Concentration is responsible for the entire peculiar motion of the Local Group (of 570 km s^{-1}), for which it would have to have a mass of $M_{\text{GA}} \approx 2.6 \pm 0.8 \times 10^{17} h^{-1} M_{\odot}$ at a distance of $136 h^{-1} \text{ Mpc}$ ($H_0 = 100 h \text{ km s}^{-1}$). If all the 3,390 galaxies belonging to the Hydra-Centaurus catalogue, within a radius of 10° of Shapley 8 ($24 h^{-1} \text{ Mpc}$ at a distance of $136 h^{-1}$) and smaller than $D_{\text{ESO}} = 1.5'$, were placed at a distance of $136 h^{-1} \text{ Mpc}$, and their magnitudes corrected for galactic extinction¹⁶, then the total luminosity would be $L_J = 4.1 \times 10^{13} h^{-2} L_{\text{B},\odot}$, leading to $M/L_{\text{B}} \approx M/L_J \approx$

TABLE 1 Clusters of galaxies found in the Hydra-Centaurus catalogue

*	No. ACO†	Position (1950)‡		$\langle v_r \rangle$ § km s^{-1}	σ km s^{-1}	N_v ¶	Other names**
		RA	Dec				
A	842	9 h 35.9 min	-20° 42'				
B		9 57.3	-31 9				
27		10 23.0	-39 31	2,725	139	8	N 3244
29		10 28.4	-34 59	2,860	546	20	Antlia, P 636
34	1060	10 34.5	-27 16	3,417	623	103	Hydra I
C	3494	11 54.6	-31 53				
D		11 57.6	-20 16				
59		12 1.9	2 11	5,850	708	13	MKW-4
E	1520	12 17.0	-13 0				
32		12 26.8	-39 10	3,194	268	6	N 4373
30	3526a	12 46.1	-41 2	3,041	586	123	Centaurus 30
46	3526b			4,570	262	57	Centaurus 45
152*	1631	12 50.2	-15 10	13,958	711	71	
F	1633	12 51.5	-26 7				
161	3528	12 51.6	-28 45	16,050	740	39	K 21
G	1635	12 51.7	-8 40				
135*	1644	12 54.6	-17 6	14,202	991	92	
H	3532	12 54.6	-30 6				K 22
50	3537	12 58.3	-32 10	5,010			K 23
I		13 7.7	-7 9				
116	3545	13 8.6	-33 49	11,550			
J		13 12.0	-16 13				
L		13 15.9	-37 0				
290	1709	13 16.0	-21 12	28,980			
46	3554	13 16.7	-33 13	4,590			
150'		13 18.7	-35 32	14,970			K 24, P 729
M	1725	13 21.2	-16 33				
147	3556	13 21.3	-31 24	14,700	590	12	
P	3557	13 22.1	-28 37				
N		13 24.1	-12 0				
101*	1736a	13 24.1	-26 52	10,412	381	34	IC 4255
	1736b			13,815	984	63	
151		13 25.0	-40 52	15,090			P 734
143	3558	13 25.1	-31 14	14,300	1,250	91	Shapley 8
138	3559	13 27.1	-29 16	13,750	600	9	
258	1750	13 28.3	-1 36	25,800		1	
33	3560	13 29.0	-32 58	3,270			
150	3562	13 30.7	-31 25	14,970			
Q	1757	13 30.8	-23 1				K 26?
R	3564	13 31.5	-34 58				
33'	3565	13 33.8	-33 43	3,270			IC 4296
S	3566	13 36.1	-35 18				
T		13 41.3	-19 33				P 741
U		13 41.7	-34 43				P 742
112	3570	13 43.9	-37 40	11,160			
117	3571	13 44.6	-32 37	11,730	1,060	69	
42	3574	13 46.3	-30 3	4,230			K 27
V	3577	13 51.5	-27 36				
W	3578	13 59.7	-24 29				
42'		14 0.7	-33 44	4,200			P 753
Y		14 1.5	-32 43				
63	3581	14 4.6	-26 47	6,270			PKS 1406-267

This table lists the galaxy clusters found in the APM Hydra-Centaurus Survey, indicated in Fig. 1b. The columns indicate: * the marker used in Fig. 1b (letters for clusters with no available redshift, numbers refer to redshift of the cluster in 100 km s^{-1} , except for those marked with an asterisk for which the redshifts were updated later); † 'Abell' numbers (from the ACO rich cluster catalogue³); ‡ position (1950); § mean redshift with respect to the Local Group, and velocity dispersion, in km s^{-1} , where available (refs 6-10, and Quintana and da Souza, preprint); ¶ the number of galaxies used in estimating $\langle V_r \rangle$ and σ , if known; and ** Alternative names found in the literature (P, ACO poor clusters³, K, Klemola²¹).

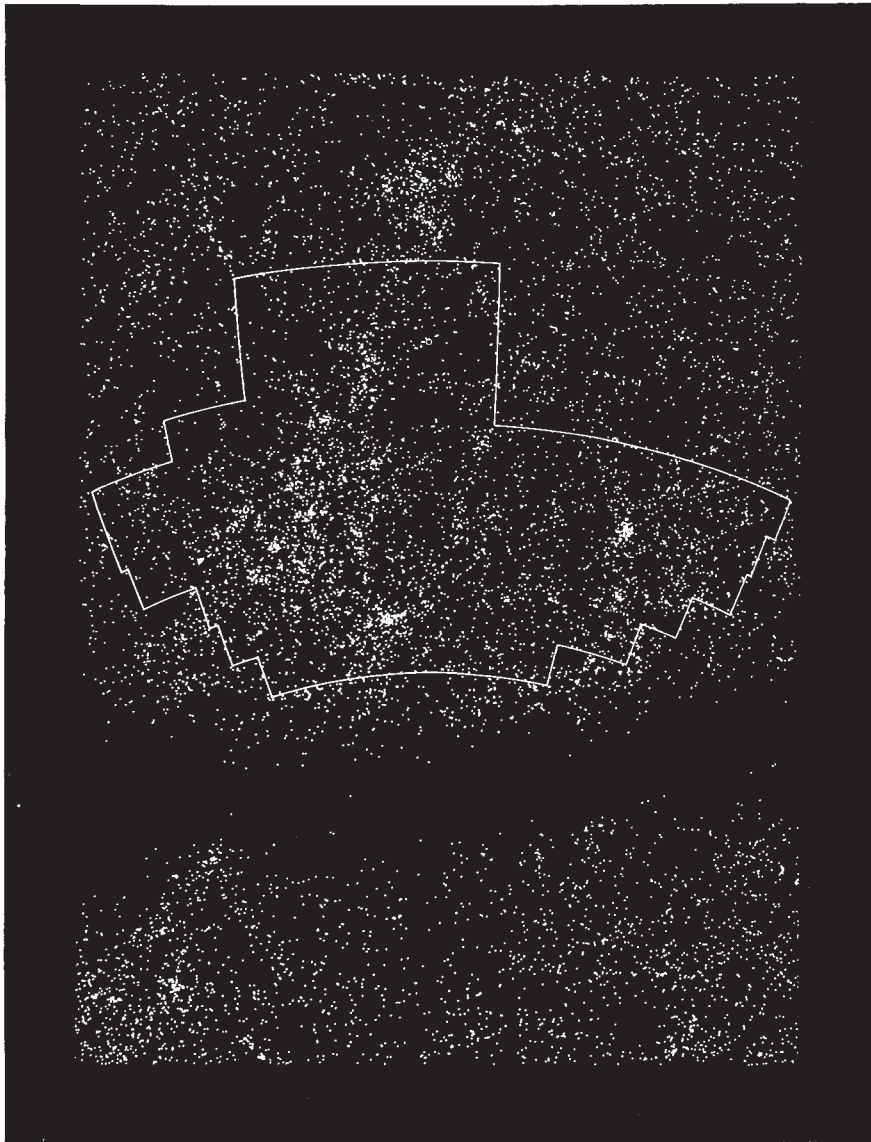


FIG. 2 An equal-area plot of all galaxies in the survey region and its immediate neighbourhood, as found in existing catalogues (a merged version of the ESO, MCG, UGC catalogues as used in Fig. 8 of ref. 1). The area enclosed by the survey region (outline indicated) contains 1,220 galaxies larger than $D_{\text{ESO}} > 1.3'$. The cluster to the north of the survey area is the Virgo Cluster. The Supergalactic Plane runs across the plot, on either side of the galactic plane, the latter showing up as an empty 'zone of avoidance'. The nearby clusters ($\langle v_r \rangle < 5,000 \text{ km s}^{-1}$, such as Hydra I and Centaurus) are easily seen, but not those belonging to the further Shapley Concentration.

$6,300 h$ (for comparison: $M/L_B \approx 1,500 h$ for an $\Omega_0 = 1$ Universe). This makes the Shapley Concentration unlikely to be the Great Attractor, as it would also imply (1) that mass does not follow light, and dark matter prefers to reside in clusters and superclusters of galaxies rather than in the field, and furthermore, (2) that superclusters are capable of having higher mass-to-light ratios ($M/L_B \approx 6,000 h$) than the richest clusters of galaxies (for example, Perseus: $M/L_B \approx 850 h$). Both of these are against biasing arguments.

Furthermore, if the Shapley Concentration were a major source of the dipole acceleration of the Local Group, one would expect peculiar velocities to continue to rise beyond the nearer concentration of galaxies. However, preliminary results¹⁸ from Dressler and Faber's survey in this region indicate that peculiar velocities of galaxies decrease beyond $4,000 \text{ km s}^{-1}$. The Hydra-Centaurus catalogue presented here can serve as a source for more extensive surveys of this kind.

Massive 'attractors' can have an appreciable effect on the cosmic microwave background radiation (CMBR). Rees and Sciama¹⁹ showed that a nearby expanding (contracting) density enhancement, such as a supercluster, will cause CMBR photons passing through it to be blueshifted (redshifted) relative to photons reaching us from other directions, causing an anisotropy in the CMBR on the scale of the size of the supercluster (a few degrees). If the entire mass of $2.7 \times 10^{16} h^{-1} M_\odot$ of the Great

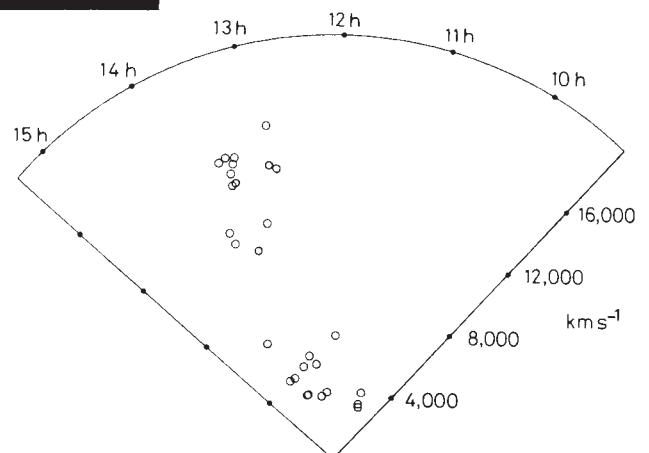


FIG. 3 A wedge diagram of galaxy clusters with known redshifts, showing the position on the sky (right ascension) against mean redshift. The plot clearly shows the existence of two distinct groups of clusters. The nearer group is known as the 'Hydra-Centaurus Supercluster', consisting of three subgroups: (1) the 'Hydra' subgroup at a distance of $\langle v \rangle = 3,000 \text{ km s}^{-1}$, (2) the 'Centaurus' subgroup at $\langle v \rangle = 3,200 \text{ km s}^{-1}$; and (3) the Klemola 27 subgroup at $\langle v \rangle = 4,500 \text{ km s}^{-1}$. The further group, between $10,000 < v < 16,500 \text{ km s}^{-1}$, is the 'Shapley Concentration', the mean redshift being $\langle v \rangle = 13,620 \text{ km s}^{-1}$. Note the absence of clusters between 5,000 and 10,000 km s^{-1} , and beyond 3,500 km s^{-1} in the range $9 \text{ h} < \text{RA} < 12 \text{ h}$ 30 min.

Attractor¹, at a distance of $4,350 \text{ km s}^{-1}$, were confined to a uniform sphere of radius $15 h^{-1} \text{ Mpc}$ ($\delta\rho/\rho_0 = 6\Omega_0^{-1}$), it would produce $|\Delta T/T| \approx 2 \times 10^{-6}$ for $\Omega_0 = 1$. On the other hand, if the Shapley Concentration at $13,600 \text{ km s}^{-1}$ were the Great Attractor, and could be represented as a spherical supercluster of mass $2.6 \times 10^{17} h^{-1} M_\odot$ within a radius of $20 h^{-1} \text{ Mpc}$ ($\delta\rho/\rho_0 = 27\Omega_0^{-1}$), it would lead to a Rees-Sciama effect of the order of $|\Delta T/T| \approx 4 \times 10^{-5}$. This is close to the observable limits of the Davies *et al.* beam-switching experiment²⁰, which in its present configuration cannot point so far south in declination.

Both the values of the luminosity and the optical (diameter) dipole of the Shapley Concentration indicate that it could be responsible for $\leq 10\%$ of the net acceleration of the Local Group, the rest being due to much nearer galaxies. It would also have a slightly greater dynamical effect on the nearer galaxy concentration, which in turn retains its status as the principal 'Great Attractor' as far as the Local Group is concerned. $\square$

Received 27 June; accepted 20 September 1989.

- Lynden-Bell, D. *et al. Astrophys. J.* **326**, 19–49 (1988).
- Scaramella, R., Baiesi-Pillastrini, G., Chincarini, G., Vettolani, G. & Zamorani, G. *Nature* **338**, 562–564 (1989).
- Abell, G. O., Corwin, H. G. & Olowin, R. P. *Astrophys. J. Suppl.* **70**, 1–138 (1989).
- Da Costa, L. N. *et al. Astr. J.* **91**, 6–12 (1986).
- Dressler, A. *Astrophys. J.* **329**, 519–526 (1988).
- Melnick, J. & Moles, M. *Rev. Mexicana Astr. Astrof.* **14**, 72–76 (1987).
- Hopp, U. & Materne, J. *Astr. Astrophys. Suppl.* **61**, 93–106 (1985).
- Lucey, J. R., Currie, M. J. & Dickens, R. J. *Mon. Not. R. astr. Soc.* **221**, 453–472 (1986).
- Metcalfe, N., Godwin, J. G. & Spenser, S. D. *Mon. Not. R. astr. Soc.* **225**, 581–590 (1987).
- Dressler, A. & Schectman, S. A., *Astr. J.* **95**, 284–297 (1988).
- Lynden-Bell, D., Lahav, O. & Burstein, D. *Mon. Not. R. astr. Soc.* (in the press).
- Shapley, H. *Harvard Coll. Obs. Bull.* **874**, 9–12 (1930).
- Lahav, O., Edge, A. C., Fabian, A. C. & Putney, A. *Mon. Not. R. astr. Soc.* **238**, 881–895 (1989).
- Lahav, O., Rowan-Robinson, M. & Lynden-Bell, D. *Mon. Not. R. astr. Soc.* **234**, 677–701 (1988).
- Lauberts, A. *The ESO/Uppsala Survey of the ESO-B Atlas* (European Southern Observatory, Garching, 1982).
- Burstein, D. & Heiles, C. *Astr. J.* **87**, 1165–1189 (1982).
- Maddox, S. J., Efsthathiou, G. & Loveday, J. in *Large Scale Structures of the Universe*, Proc. IAU 130 (eds Audouze, J., Pelletan, M.-C. & Szalay, A.) 151–160 (Kluwer, Dordrecht 1987).
- Dressler, A. in *Proc. 14th Texas Symp. Relativistic Astrophys.* (New York Acad. Sci., in the press).
- Rees, M. J. & Sciama, D. W. *Nature* **217**, 511–516 (1968).
- Davies, R. D. *et al. Nature* **326**, 462–465 (1987).
- Klemola, A. R. *Astr. J.* **74**, 804–806 (1969).

ACKNOWLEDGEMENTS. I thank the members of the APM group at Cambridge for scanning the UKST plates, in particular M. J. Irwin, S. J. Maddox and J. Loveday for their APM reduction software. I am grateful to D. Lynden-Bell for advice and encouragement, to J. Loveday for supplying the values of (Δ_2^2) , to M. J. Rees and A. C. Fabian for discussions and to S. Tritton and C. Thoburn for supplying plates at short notice. The author is supported by the Isaac Newton Studentship at the University of Cambridge.

Pre-biotic organic matter from comets and asteroids

Edward Anders

Enrico Fermi Institute and Department of Chemistry, University of Chicago, Chicago, Illinois 60637-1433, USA

SEVERAL authors^{1–3} have suggested that comets or carbonaceous asteroids contributed large amounts of organic matter to the primitive Earth, and thus possibly played a vital role in the origin of life. But organic matter cannot survive the extremely high temperatures ($>10^4 \text{ K}$) reached on impact, which atomize the projectile and break all chemical bonds. Only fragments small enough to be gently decelerated by the atmosphere—principally meteors of 10^{-12} – 10^{-6} g —can deliver their organic matter intact⁴. The amount of such 'soft-landed' organic carbon can be estimated from data for the infall rate of meteoritic matter. At present rates, only $\sim 0.006 \text{ g cm}^{-2}$ intact organic carbon would accumulate in 10^8 yr , but at the higher rates of $\sim 4 \times 10^9 \text{ yr}$ ago, about 20 g cm^{-2} may have accumulated in the few hundred million years between the last cataclysmic impact and the beginning of life. It may have included some biologically important compounds that did not form by abiotic synthesis on Earth.

Figure 1 shows the present estimates of the meteoritic mass influx on Earth⁵, including meteors, meteorites and crater-forming bodies (asteroids and comets). The meteor data⁶ cover only the past few decades, but as they agree⁷ with noble-metal measurements in deep-sea sediments^{5,8} and lunar soils⁹, they should be representative of the past $\sim 3.6 \times 10^9 \text{ yr}$. The cratering curve should be similarly representative, being based in part on crater counts on lunar maria that are 3.3 – $3.8 \times 10^9 \text{ yr}$ old (ref. 10).

Now consider the size ranges in which organic matter can be preserved (shaded areas in Fig. 1). At very small masses, ultraviolet photolysis alters and destroys the organic matter. For the sake of definiteness I set the limit for such destruction at $<10^{-12} \text{ g}$. At higher masses, atmospheric friction heats the particle to increasingly higher temperatures. The nominal threshold for boiling of silicates is $>10^{-6} \text{ g}$, but it depends strongly on density, velocity, entry angle and fragmentation^{6,11}. Well preserved interplanetary dust particles (containing noble gases, nuclear tracks and organic matter) range up to $50 \mu\text{m}$ in size¹¹ and as many of these may be fragments of larger aggregates or particles, I adopt 10^{-6} g as the upper limit for survival of organic matter. In the next few intervals, much or all of the mass is lost by ablation^{6,11}, and any remainder is strongly heated by conduction. Ablation calculations and thermoluminescence data show that pyrolytic temperatures ($>600^\circ\text{C}$) generally extend $<1 \text{ mm}$ into the meteorite¹², and thus survival of organic matter—at least for low-velocity objects—becomes possible at larger masses, say 10 g . At still larger masses, atmospheric deceleration becomes progressively less effective, causing increasing destruction in the atmosphere¹³ or on impact. A reasonable upper limit is 10^8 g , as meteorites of that mass usually reach the ground with a significant fraction of their cosmic velocity. (Also, the largest stony meteorite showers have masses of only a few tonnes.)

It is useful to estimate the accretion rate not only of organic C but also of total C. The first four lines of Table 1 give such data for the main types of meteoritic material. A few numbers require justification.

For meteors, I use the mean carbon value of $\sim 10\%$ for interplanetary dust particles¹⁴ rather than the 24% reported for comet Halley dust¹⁵, as the latter was measured near the comet nucleus where ices had not yet completely vapourized. This value should also apply to dead comets, which are far more likely to strike the Earth than live comets. (The luminous lifetime of an Earth-crossing short-period comet is typically $\sim 10^3$ – 10^4 revolutions, whereas its dynamical lifetime is $\sim 10^7$ revolutions.) For the crater-forming bodies, I assume equal numbers of comets, carbonaceous asteroids and non-carbonaceous asteroids. Several lines of evidence indicate a ratio of $\sim 1:1$ for the two asteroid classes (see below), but the fraction of comets is poorly constrained, with estimates^{10,16} in the range 10 – 70% . For meteorites, we assume a carbon content of 1.3% , representing 50% ordinary chondrites (median $\sim 0.1\% \text{ C}$) and 50% carbonaceous chondrites (range 0.35 – $4.8\% \text{ C}$). Although the latter constitute only 6% of observed stony meteorite falls, this rarity reflects preferential destruction in the atmosphere owing to their friable nature. Carbonaceous objects seem to account for about one-half of the photographic fireballs¹⁷ with final velocities $\leq 8 \text{ km s}^{-1}$, one-half of the asteroids near the 3:1 Kirkwood gap which are thought to be the principal source of meteorites¹⁸, 13 out of 32 Apollo-Amor asteroids¹⁹, and 3 out of 5 chondritic impactors chemically identified at terrestrial craters²⁰.

Next, I consider the unmelted material. Meteors pass through the atmosphere intact, but meteorites lose much of their mass by ablation. The ablation loss depends strongly on velocity, mass and crushing strength¹³. I shall use 90% for non-carbonaceous chondrites²¹ and 99.4% for carbonaceous chondrites (the latter figure is chosen to reduce the proportion of C-chondrites from 50% to 6%).

Obviously, meteors supply most of the organic carbon. It will not be pristine, having been altered by solar ultraviolet and

atmospheric heating, but this could make its chemistry more, rather than less, interesting. Meteorites supply only a negligible fraction ($\sim 2 \times 10^{-5}$) of the total.

The accretion rate of intact organic C, $3.2 \times 10^8 \text{ g yr}^{-1}$, would give $3.2 \times 10^{16} \text{ g}$, or 0.006 g cm^{-2} , in 10^8 yr . This amount is either large or small, depending on the frame of reference: although five times greater than the mass of the marine biosphere, it corresponds to a mean carbon concentration in the oceans of only $2 \times 10^{-5} \text{ g l}^{-1}$. But the true value must be higher, because meteorite infall rates were much higher during the first $\sim 10^9 \text{ yr}$ of the Earth's history, when the inner planets swept up left-over planetesimals as well as asteroids and comets scattered by the giant planets. To assess the amount of organic carbon contributed by this 'early intense bombardment', we first consider the origin of the Earth's crustal carbon ($1.4 \times 10^{23} \text{ g}$; ref. 22).

Less than 10^{-3} of this carbon can come from meteorite infall at present rates over $4.5 \times 10^9 \text{ yr}$. Even if we count all meteors in Fig. 1 and integrate the comet-asteroid distribution to $m = 8.7 \times 10^{20} \text{ g}$ (one impact of this mass or larger is expected in $4.5 \times 10^9 \text{ yr}$), the total mass is only $2.1 \times 10^{21} \text{ g}$ (including $7.4 \times 10^{19} \text{ g}$ meteors). For a bombarding population composed equally of comets, carbonaceous asteroids and non-carbonaceous asteroids, the mean C content is 4.2%, which gives a total carbon accretion of $9 \times 10^{19} \text{ g}$ in $4.5 \times 10^9 \text{ yr}$, or only 7×10^{-4} of the Earth's crustal C.

There is good reason to believe that the remaining C (and N, H) was acquired as a 'late veneer' of carbonaceous material during the final stages of the Earth's accretion²³⁻²⁵. Not only C, N and H but also other volatile elements (Ti, Pb, B, In, Bi, Cl, Br, I, Ar and Kr) are present in the Earth's crust in approximately CI-chondrite proportions^{23,25}. As these elements differ in volatility and atomic weight, it is unlikely that the observed, flat abundance pattern (Fig. 4 of ref. 25) could have arisen from 'impact volatilization'^{26,27}—a popular but dubious concept for the final stages of Earth's accretion. It is more likely to reflect an original nebular condensation pattern that was preserved during formation of the crust. (These elements are incompatible with mantle minerals and would strongly concentrate in the crust during mantle differentiation.) The threshold of $\sim 400 \text{ K}$ for carbon condensation as complex organic matter²⁸ is still recognizable in the asteroid belt, where dark, carbonaceous asteroids become dominant at $> 2.5 \text{ AU}$ (ref. 29). About $4.0 \times 10^{24} \text{ g}$ of CI-chondrite-like material ($6.7 \times 10^{-4} M_{\oplus}$, or a 3.5-km layer) would suffice to account for the $1.4 \times 10^{23} \text{ g}$ carbon at the Earth's surface, and the amounts needed for N and H are rather similar (2.9×10^{-4} and $1.5 \times 10^{-3} M_{\oplus}$).

Two further lines of evidence indicate an infall of this order. First, the uniform enrichment of noble metals (Pd, Re, Os, Ir, Au) in the upper mantle, at $(7.32 \pm 0.99) \times 10^{-3}$ CI-chondrite abundances³⁰, apparently implies that highly oxidized, chon-

dritic material fell on the Earth after core formation, leaving its siderophile elements stranded^{30,31}. One mantle xenolith showing such enrichment, a sheared garnet lherzolite of particularly primitive composition³², comes from $> 150 \text{ km}$ depth and so the 'megaregolith' from the late bombardment presumably extends to at least this depth. This would imply $> 5 \times 10^{-4} M_{\oplus}$ of chondritic material in the upper mantle, which is remarkably consistent with the above values for C, N, H and noble metals (overall mean = $(7.4 \pm 2.7) \times 10^{-4} M_{\oplus}$). Second, the size and age distribution of impact basins on the Moon indicates that the Earth accreted 1.3×10^{-4} – $3.8 \times 10^{-2} M_{\oplus}$ of material between 3.8 and 4.5 Gyr ago³³. Some authors have postulated a very large influx of cometary bodies from the region of the outer planets³⁴, but as only one of the last nine large bodies to strike the Moon^{35,36} resembled C-chondrites³⁰, only part of the bombarding population was carbonaceous.

Although this late accretional component was $\sim 10^3$ times greater than the total meteoritic influx at present rates, it probably yielded a smaller proportion of intact organic matter. Like today, the most efficient vehicle for delivery of organic matter would be meteors in the range 10^{-12} – 10^{-6} g (or higher if the atmosphere was denser). The meteor fraction probably was at least as large as the present one, that is, $\sim 20\%$ of the total (Table 1), as the disintegration rate of comets should have been similar and the collision rate in the asteroid belt was higher³⁷. Much of the soft-landed organic matter would be destroyed or degraded, however, first by the high temperatures of the crust and atmosphere during the most intense stages of the bombardment, then by recurrent vapourization of the ocean and by burial. Material cycled through the $> 150\text{-km}$ megaregolith would be pyrolysed to elemental C and small molecules such as CO and CH_4 , so only organics accumulated in the ocean and in the last few kilometres of the regolith would escape serious degradation.

A reasonable guess—good to perhaps a factor of 3–5—is that only $\sim 1\%$ of the Earth's surface carbon ($\sim 1.4 \times 10^{21} \text{ g}$) accreted under conditions permitting survival of organic carbon. (One per cent of the accretable material ($\sim 5 \times 10^{22} \text{ g}$) would suffice to evaporate the ocean and to deposit several kilometres of ejecta, so accumulation of organic matter probably began only during the last few per cent of the accretion process, but at first with very low efficiency.) If the fraction of meteors, and hence intact organic carbon, was the same as in the most recent infall, that is $\sim 8\%$ (Table 1), this corresponds to $1.1 \times 10^{20} \text{ g}$ organic C, or 10^2 times the total infall in $4.5 \times 10^9 \text{ yr}$ at present rates. This amount is equivalent to 21 g cm^{-2} or an oceanic carbon concentration of 0.08 g l^{-1} .

Even $1.1 \times 10^{20} \text{ g}$ C is comparable to the 10^{20} – 10^{21} g of HCN and HCHO expected from Miller-Urey reactions in the Earth's atmosphere ($\text{H}_2/\text{CO}_2 = 1$) over 600 Myr (ref. 38). Amounts aside, meteoritic and cometary organic matter may well have

TABLE 1 Accretion rates on Earth

Sources	Mass range (g)	Mass accretion rate (Gg yr^{-1})	Carbon content (%)	Carbon accretion rate (Gg yr^{-1})
Total meteoritic matter	10^{-14} – 10^{18}	78	—	4.2
Meteors	10^{-14} – 10^2	16	10	1.6
Meteorites*	10^1 – 10^8	0.058	1.3	7.6×10^{-4}
Crater-forming bodies†	10^8 – 10^{18}	62	4.2	2.6
Unmelted material, contributing organic matter				
Meteors	10^{-12} – 10^{-6}	3.2	10	0.32
Meteorites, noncarbonaceous‡	10^1 – 10^8	2.9×10^{-3}	0.1	2.9×10^{-6}
Meteorites, carbonaceous§	10^1 – 10^8	1.9×10^{-4}	2.5	4.7×10^{-6}

* Assumed pre-atmospheric abundances: 50% carbonaceous chondrites (2.5% C) and 50% noncarbonaceous meteorites (0.1% C).

† Assuming equal mass fractions of dead comets (10% C), carbonaceous asteroids (2.5% C), and noncarbonaceous asteroids (0.1% C).

‡ Assumed ablation loss 90%.

§ Assumed ablation loss 99.4%, to yield observed fraction of 6% carbonaceous chondrites among stony meteorites.

played an important role in pre-biotic evolution, as it formed by different processes (ion-molecule and photochemical reactions in interstellar clouds, grain catalysis in the solar nebula, aqueous chemistry in asteroids) and thus contains a different mix of compounds, including N-heterocyclics, aromatics, amphiphilic compounds, porphyrin-like pigments, and so on, that are not readily made by Miller-Urey reactions^{28,39-41}.

There are several uncertainties in this estimate: the fraction of carbon-rich objects in the final stages of bombardment, the fraction disintegrated to meteor-sized dust, the onset of safe conditions on Earth, the pressure of the atmosphere, and the net yield of organics from large, friable bodies. A higher pressure would extend aerobraking to higher masses⁴², thus raising the soft-landed fraction in Fig. 1. Survival of dust from large, friable bodies would have the same effect. Öpik⁴³ has calculated that aerodynamic pressure would cause large stony meteorites to disintegrate into a flat sheet of dust and rubble, which would be decelerated to less than crater-forming velocities at masses $\leq 10^{12}$ g. Thomas *et al.*⁴² have suggested that part of this debris might be decelerated so gently as to permit survival of organic matter.

At first sight, the empirical data are permissive, or even encouraging. No unmelted material has been found from the Tunguska object, but this may merely reflect its inconspicuous nature and lack of ferromagnetism. Very large amounts of two non-protein amino acids (AIB and ISOVAL) have been reported $+0.3$ m and -0.5 m from the Cretaceous/Tertiary (K/T) boundary at Stevns Klint, and have been attributed to soft-landed material from a comet⁴⁴. However, these results are in sharp conflict with measurements of Xe in 9 K/T boundary samples from 3 sites, including 6 from Stevns Klint^{45,46}. The AIB/Ir ratio is about five times the value in the Murchison C2-chondrite and implies essentially complete survival of the impactor, whereas the Xe/Ir values range from $\leq 6 \times 10^{-3}$ to 2×10^{-5} of the C-chondrite value, and imply essentially complete destruction. Differences in temperature sensitivity of AIB and Xe cannot explain this discrepancy, as amino acids are far more readily lost than is Xe (seconds at 1,000 K against hours at 1,200–1,500 K), and neither can differences in composition. Comets are ~ 10 -fold enriched in C and N and hence perhaps in amino acids, but they are not likely to be significantly depleted in Ir, as judged from the C-chondrite-like pattern of the micro-

meteorite component in lunar soils⁹ and from the relatively normal abundances of cosmochemically similar elements in comet Halley¹⁵ (Fe = 0.62, Ni = 0.87, Ca = 1.10 of C1-chondrite values). Xenon actually is more abundant in IDPs⁴⁷ than in C1 chondrites, by a factor of about ten.

Furthermore, there are theoretical arguments against survival of unmelted material from comets the size of the K/T impactor ($\sim 10^{18}$ g). During atmospheric passage, such a body would lose only $\sim 10^{-3}$ of its momentum or mass, unless it had a very small entry angle or very irregular (non-equant) shape⁴⁸. Moreover, the 'ground track' a of the comet's atmospheric trajectory is $h/\tan \alpha$, where h is the altitude where significant fragmentation begins and α is the elevation angle. If $h = 50$ km (ref. 49) and $\alpha \geq 20^\circ$, then a is only ≤ 137 km, much smaller than the $\sim 1,200$ -km radius of the fireball. Any debris spalled from the comet during atmospheric passage will be engulfed by the fireball and destroyed.

Survival of organic matter against this destruction mechanism becomes possible when the fireball radius is shorter than the ground track, that is at masses $< 10^{15}$ g. If a significant fraction of the spalled-off dust is gently decelerated⁴² without melting or pyrolysis, then friable bodies of 10^8 – 10^{12} g may contribute some organic matter. However, at present it seems that the best vehicle for carrying organic matter to Earth is cometary dust formed before atmospheric entry. □

Received 7 July; accepted 6 October 1989.

- Oró, J. *Nature* **190**, 389–390 (1961).
- Delsemme, A. H. *Origins Life* **14**, 51–60 (1984).
- Deamer, D. W., Brack, A., Morowitz, H., Weber, A. & Usher, D. *Origins Life* (in the press).
- Anders, E., Hayatsu, R. & Studier, M. H. *Science* **182**, 781–790 (1973).
- Kyte, F. T. & Wasson, J. T. *Science* **232**, 1225–1229 (1986).
- Hughes, D. W. in *Cosmic Dust* (ed. McDonnell, J. A. M.) 123–185 (Wiley, New York, 1978).
- Dohnanyi, J. S. *Science* **173**, 558 (1971).
- Barker, J. L. & Anders, E. *Geochim. cosmochim. Acta* **32**, 627–645 (1968).
- Anders, E., Ganapathy, R., Krähenbühl, U. & Morgan, J. W. *The Moon* **8**, 3–24 (1973).
- Shoemaker, E. M. A. *Rev. Earth planet. Sci.* **11**, 461–494 (1983).
- Brownlee, D. E. A. *Rev. Earth planet. Sci.* **13**, 189–215 (1985).
- Sears, D. W. *Mod. Geol.* **5**, 155–164 (1975).
- Baldwin, B. & Shaeffer, Y. *J. geophys. Res.* **76**, 4653–4668 (1971).
- Schramm, L. S., Brownlee, D. E. & Wheelock, M. M. *Meteoritics* **24**, 99–112 (1989).
- Jessberger, E. K., Christoforidis, A. & Kisseil, J. *Nature* **332**, 691–695 (1988).
- Weissman, P. R. *Geol. Soc. Am. Spec. Pap.* **190**, 15–24 (1982).
- Wetherill, G. W. & ReVelle, D. O. *Icarus* **48**, 308–328 (1981).
- Wetherill, G. W. *Phil. Trans. R. Soc. A* **323**, 323–337 (1987).
- Wetherill, G. W. & Shoemaker, E. M. *Geol. Soc. Am. Spec. Pap.* **190**, 1–13 (1982).
- Palme, H., Grieve, R. A. F. & Wolf, R. *Geochim. cosmochim. Acta* **45**, 2417–2424 (1981).
- ReVelle, D. O. *J. Atmos. terr. Phys.* **41**, 453–473 (1979).
- Mason, B. *Principles of Geochemistry* 3rd edn (Wiley, New York, 1966).
- Anders, E. *Acc. chem. Res.* **1**, 289–298 (1968).
- Turekian, K. K. & Clark, S. P. Jr. *Earth planet. Sci. Lett.* **6**, 346–348 (1969).
- Anders, E. & Owen, T. *Science* **198**, 453–465 (1977).
- Matsui, T. & Abe, Y. *Nature* **322**, 526–528 (1986).
- Zahnle, K. J., Kasting, J. F. & Pollack, J. B. *Icarus* **74**, 62–97 (1988).
- Hayatsu, R. & Anders, E. *Topics curr. Chem.* **99**, 1–37 (1981).
- Grady, J. & Tedesco, E. *Science* **216**, 1405–1407 (1982).
- Morgan, J. W. *J. geophys. Res.* **91**, 12375–12387 (1986).
- Kimura, K., Lewis, R. S. & Anders, E. *Geochim. cosmochim. Acta* **38**, 683–701 (1974).
- Morgan, J. W., Wandless, G. A., Petrie, R. K. & Irving, A. J. *Tectonophysics* **75**, 47–67 (1981).
- Chyba, C. F. *Nature* **330**, 632–635 (1987).
- Ip, W.-H. & Fernandez, J. A. *Icarus* **74**, 47–61 (1988).
- Gros, J., Takahashi, H., Hertogen, J., Morgan, J. W. & Anders, E. *Proc. Lunar Sci. Conf.* **7**, 2403–2435 (1976).
- Hertogen, J., Janssens M.-J., Takahashi, H., Palme, H. & Anders, E. *Proc. Lunar Sci. Conf.* **8**, 17–45 (1977).
- Chapman, C. R. & Davis, D. R. *Science* **190**, 553–556 (1975).
- Stribling, S. & Miller, S. L. *Origins Life* **17**, 261–273 (1987).
- Deamer, D. W. *Nature* **317**, 792–794 (1985).
- Mullie, F. & Reisse, J. *Topics curr. Chem.* **139**, 83–117 (1987).
- Cronin, J. R., Pizzarello, S. & Cruikshank, D. P. in *Meteorites and the Early Solar System* (eds Kerridge, J. F. & Matthews, M. S.) 819–857 (Univ. Arizona Press, 1988).
- Thomas, P. J., Chyba, C. F., Brookshaw, L. & Sagan, C. *Lunar planet. Sci.* **20**, 1117–1118 (1989).
- Öpik, E. J. in *Proc. Geophys. Lab. Lawrence Radiation Lab. Cratering Symp.*, Rep. UCRL-6438, **2**, Paper S. 1–28 (Lawrence Radiation Lab., Berkeley, 1961).
- Zhao, M. & Bada, J. L. *Nature* **339**, 463–465 (1989).
- Wolbach, W. S., Lewis, R. S. & Anders, E. *Science* **230**, 167–170 (1985).
- Lewis, R. S. & Wolbach, W. S. *Meteoritics* **21**, 434–435 (1986).
- Hudson, B., Flynn, G. J., Fraundorf, P., Hohenberg, C. M. & Shirck, J. *Science* **211**, 383–386 (1981).
- Emiliani, C., Kraus, E. B. & Shoemaker, E. M. *Earth planet. Sci. Lett.* **55**, 317–334 (1981).
- Turco, R. P. *et al.* *Icarus* **50**, 1–52 (1982).

ACKNOWLEDGEMENT. I thank G. W. Wetherill for advice. This work was supported in part by NASA.

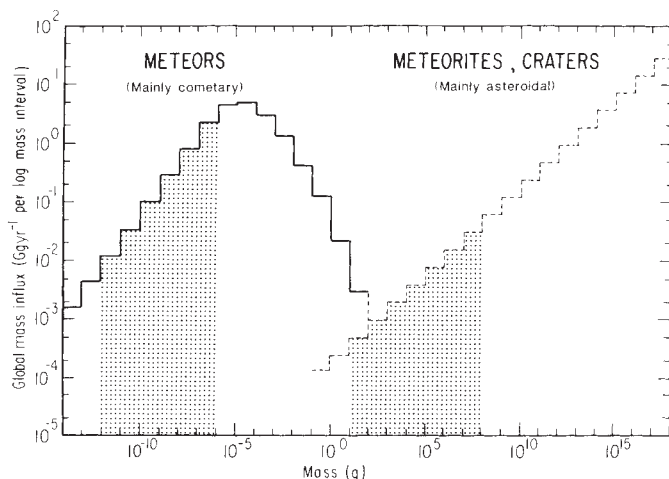


FIG. 1 Infall rate of meteoritic matter on Earth (adapted from ref. 5). Intervals where organic matter can survive passage through atmosphere are shaded. The curve on the right is based on the relation⁵ $N = 0.54 r^{-2.1}$ (N = number of impacts per Myr, r = radius in km), for an assumed density of 3 g cm^{-3} . The corresponding mass accretion rate (Gg yr^{-1}) between r_1 and r_2 is $15.83 (r_2^{0.9} - r_1^{0.9})$.

Demonstration of the tunnel-diode effect on an atomic scale

P. Bedrossian^{*†}, D. M. Chen^{*†‡}, K. Mortensen^{*†}
& J. A. Golovchenko^{*†}

^{*} Lyman Laboratory, Harvard University, Cambridge, Massachusetts 02138, USA

[†] Rowland Institute for Science, Cambridge, Massachusetts 02142, USA

[‡] Physics Department, University of Western Ontario, London, Ontario, Canada, N6A 3K7

THE tunnel diode¹, which is widely used in high-speed electronics applications², depends on the property of negative differential conductivity, that is, a negative slope in the current–voltage curve. The mechanism underlying the tunnel diode's behaviour, namely the existence of a range of biases for which tunnelling is forbidden or suppressed following a bias for which tunnelling is strongly favoured, has been employed subsequently in the design of new devices that also display the conductance anomaly, such as the double-barrier resonant-tunnelling device³. It has been predicted⁴ that the conductance anomaly could result from a similar mechanism at the tunnel junction of the scanning tunnelling microscope (STM), where localized states on adsorbate and tip atoms give rise to allowed and suppressed energies for tunnelling. The STM has the capability to image regions of negative differential conductivity induced by individual atoms on a surface. Here we report the observation of negative differential conductivity on particular binding sites of a Si (111) surface doped with boron. Specific

current–voltage characteristics are shown to be related to the presence or absence of the dopant at individual atomic sites, and negative differential conductivity is observed at -1.4 V tip bias at a specific type of site. Tunnelling spectroscopy indicates that the effect results from a tunnel-diode mechanism.

Samples were prepared by annealing degenerately boron-doped Si (111) crystals (10^{20} cm⁻³) at $1,000$ °C in an ultra-high-vacuum chamber with a base pressure of 10^{-10} torr. This procedure results in boron segregation to the surface⁵, with the appearance of the $(\sqrt{3} \times \sqrt{3})R30^\circ$ surface periodicity in both low-energy electron diffraction (LEED) and STM. This structure has been established⁶ for surface boron concentrations up to one-third of a monolayer (1 monolayer = 7.8×10^{14} atoms cm⁻²) although it is not characteristic of the pure crystal, where the 7×7 reconstruction is the more stable⁷.

Figure 1a shows a typical tunnelling image of the boron-doped surface, with individual adatoms spaced 6.7 Å apart. Two types of adatom sites, apparently randomly distributed, are distinguished by their relative brightness in the tunnelling image, and this preparation typically results in the incorporation of 15–20% 'bright' sites in the surface. We have previously established (ref. 8; see also refs 9, 10) that all imaged adatoms are silicon, but with the dark adatom sites sitting directly above a substitutional boron atom and the bright adatom sites above a silicon atom. The surface boron concentration for such a sample, then, is within 20% of saturation at one-third of a monolayer. We shall refer to the lateral locations of the two types of silicon adatom sites as boron-occupied and boron-free positions, respectively. Similar results were obtained after ion-beam sputter deposition of boron onto clean, arsenic-doped (0.02Ω cm) Si (111), with the surface boron concentration controlled in that case by adjusting the deposition time.

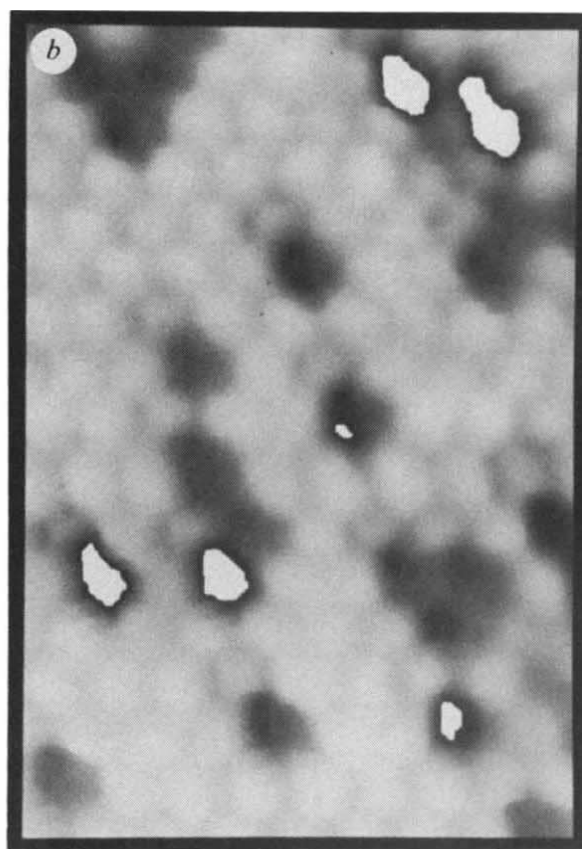
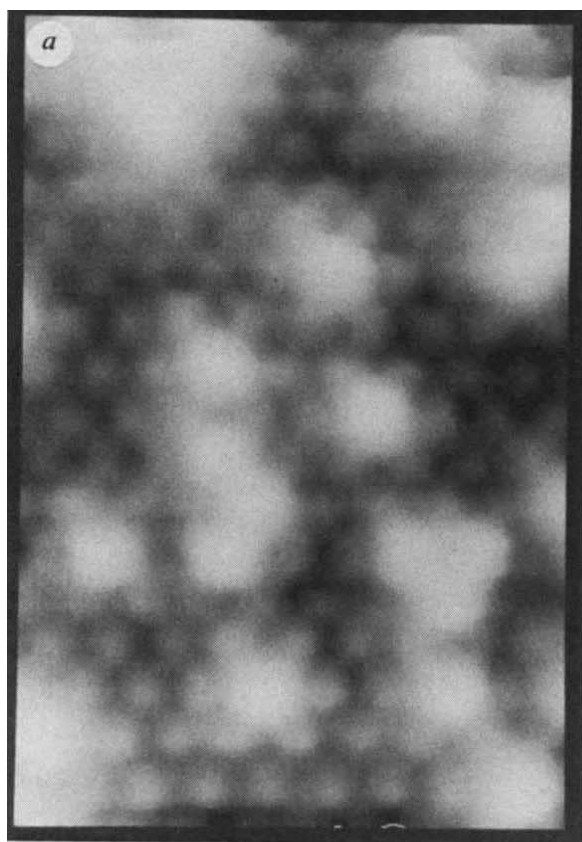


FIG. 1 a, 60×170 Å topographical tunnelling image of the Si(111) $(\sqrt{3} \times \sqrt{3})R30^\circ$ -B surface prepared by annealing heavily boron-doped Si (111). Tip bias, -1.4 V, 1 nA. The protrusions represent individual atoms. b, Image of

the differential conductivity at -1.4 V for the region in a. Brighter shading indicates higher differential conductivity. The white, highlighted regions indicate negative differential conductivity.

Having identified the two types of adatom sites observed, we can measure atom-resolved current-voltage (I - V) characteristics by sweeping the tip bias and recording the current at the crystal for a particular lateral tip position and fixed tip-sample separation. Many boron-free positions that are immediately adjacent to several boron-occupied positions exhibit negative differential conductivity at a tip-to-sample bias of -1.4 V, as shown in Fig. 2. The conductance anomaly is stable and repeatable, both for individual atomic sites and from one sample to the next. No such feature is observed for boron-occupied positions, or in large regions of boron-free sites.

The spatial distribution of the regions of negative differential conductivity is displayed in an image of the differential conductivity at -1.4 V, which we obtained by placing a 6.25 kHz, 20-mV r.m.s. modulation on the tip bias and recording the output of a lock-in amplifier that senses the sample current signal as a function of lateral position. The modulation frequency is placed outside the bandwidth of the feedback network that adjusts the vertical position of the tip and maintains a constant d.c. tunnelling current. Figure 1b displays such an image, corresponding to the region of the surface shown in the topographical image of Fig. 1a. In Fig. 1b, brighter shading indicates higher differential conductivity, and the solid white areas indicate negative differential conductivity. Comparison with Fig. 1a shows that regions of negative differential conductivity are localized on isolated boron-free sites and therefore act as individual tunnel diodes in concert with the STM tip. We note that the surface preparation that gave rise to this behaviour typically results in an average of at least one tunnel diode in a 60×60 Å region.

The mechanism and conditions behind the appearance of the conductance anomaly in this system are indicated by local tunnelling spectra, which are obtained by positioning the tip over an adatom site, fixing the lateral and vertical tip positions momentarily, sweeping the tip bias, and recording the differential conductivity as a function of this bias. When the resulting data are normalized to the total conductivity, as in Fig. 3, peaks in such spectra indicate¹¹ the presence of a surface state at the energy corresponding to the bias voltage. Negative tip bias corresponds to tunnelling into unoccupied sample states. Thus, an unoccupied state 0.6 eV above the Fermi level is found on boron-free sites but is absent from the boron-occupied positions. The occurrence of such a state at this energy is consistent with the presence of a dangling orbital on Si adatoms that was predicted¹² for clean Si (111) ($\sqrt{3} \times \sqrt{3}$)R30° and which has been observed¹³ on Si (111) 7×7 . Above this bias, the local density of states on boron-free sites falls smoothly for 0.8 eV.

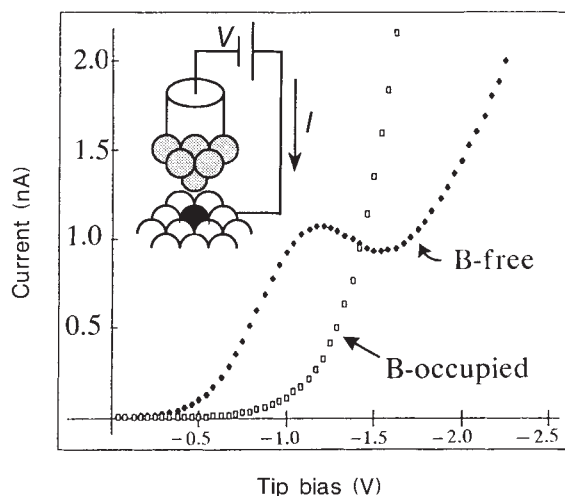


FIG. 2 I - V curves for the case of negative tip bias, acquired with the tip over boron-occupied and boron-free sites in regions of substantial mixing of the two types of sites. Inset, geometry of the tunnelling device.

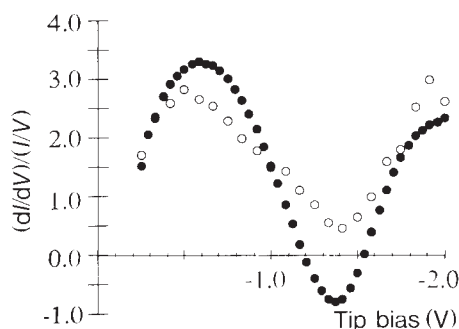


FIG. 3 Tunnelling spectra for the case of negative tip bias, acquired with the tip held over (curve a), a boron-free site surrounded by boron-free sites, and (curve b), an isolated boron-free site. The presence of the dopant leads to the negative region in curve b.

The onset of negative differential conductivity in both the tunnel diode and resonant tunnelling devices has been shown to result from a sharp drop in the tunnelling probability when the bias reaches an energy level at which tunnelling is forbidden. In the former case the forbidden energies lie in the band gap of a degenerate semiconductor; in the latter, between the discretely quantized energy levels in a potential well. In the case we consider, Fig. 3 shows that the differential conductivity at -1.4 V tip bias is low, but non-negative, even in the absence of boron. We expect the lateral isolation of a boron-free site surrounded predominantly by boron-occupied positions to narrow the dispersion of the low-lying silicon dangling-bond state, further reducing the tunnelling probability at -1.4 V bias and meeting the criterion for the conductance anomaly. The tunnel-diode-like I - V curve in Fig. 2 therefore results from a small perturbation in the surface band structure of clean Si, induced by the presence of dopant atoms below adjacent sites.

The mechanism suggested by the above data for the conductance anomaly that we observe should be distinguished from switching behaviour at electron trapping sites, which was shown¹⁴ to be responsible for time-averaged negative resistance at a variety of bias conditions in an STM study of a different system that did not display a unique dangling-bond defect. For that mechanism, switching noise in the tunnelling current was reported¹⁵ within several nanometres of the trapping site. We have not detected similar current fluctuations in the Si(111) ($\sqrt{3} \times \sqrt{3}$)R30°-B system, and we do not expect them if the conductance anomaly arises from the tunnel-diode mechanism.

Negative resistance itself can undoubtedly be realized in the STM with other systems under varying degrees of characterization and controllability of the tip and sample. Our data indicate that a particular surface can be characterized sufficiently by STM to apply a principle known to underlie the operation of a macroscopic device, specifically the tunnel diode, to engineer I - V characteristics of the tunnel junction of the STM on an atomic scale. The surface states responsible for the I - V characteristics and the nature of the atomic sites at the surface could be identified, and the overall density of tunnel diodes on the surface could be controlled, by adjusting the surface dopant concentration. Although the local dopant distribution resulting from our preparations is a product of random diffusion processes, it can be determined and selected under the STM. The recent achievement¹⁶ of large-scale integration (LSI) of STMs on a Si wafer therefore offers the means to incorporate the microscopic tunnel diode described here in a practical device. □

Received 18 April; accepted 17 October 1989.

- Esaki, L. *Phys. Rev.* **109**, 603-604 (1958).
- Sze, S. M. *The Physics of Semiconductor Devices* Ch. 9 (Wiley, New York, 1981).
- Esaki, L. *Rev. Mod. Phys.* **36**, 237-244 (1974).

4. Lang, N. D. *Phys. Rev.* **B34**, 5947-5950 (1986).
5. Korobtsov, V. V., Lifshits, V. G. & Zotov, V. G. *Surf. Sci.* **195**, 466-474 (1988).
6. Hirayama, H., Tatsumi, T. & Aizaki, N. *Surf. Sci.* **193**, L47-L52 (1988).
7. Takayanagi, K., Tanishiro, Y., Takahashi, M. & Takahashi, S. *J. Vac. Sci. Technol.* **A3**, 1502-1506 (1985).
8. Bedrossian, P. *et al.* *Phys. Rev. Lett.* **63**, 1257-1260 (1989).
9. Headrick, R. L., Robinson, I. K., Vlieg, E. & Feldman, L. C. *Phys. Rev. Lett.* **63**, 1253-1256 (1989).
10. Lyo, I.-W., Kaxiras, E. & Avouris, P. *Phys. Rev. Lett.* **63**, 1261-1264 (1989).
11. Feenstra, R., Stroscio, J. & Fein, A. *Surf. Sci.* **181**, 295-306 (1987).
12. Northrup, J. *Phys. Rev. Lett.* **57**, 154-157 (1986).
13. Nicholls, J. & Reihl, B. *Phys. Rev.* **B36**, 8071-8074 (1987).
14. Hamers, R. & Koch, R. in *The Physics and Chemistry of SiO₂ and the Si-SiO₂ Interface* (eds Helms, C. & Deal, B.) 201-210 (Plenum, New York, 1988).
15. Welland, M. & Koch, R. *Appl. Phys. Lett.* **48**, 724-726 (1986).
16. Petit, C. *Mosaic* **20**, 24-35 (1989).

ACKNOWLEDGEMENTS. This work was supported by the Office of Naval Research and the Joint Services Electronics Program.

Determination of 4-connected framework crystal structures by simulated annealing

M. W. Deem & J. M. Newsam

Exxon Research and Engineering Company, Route 22 East, Annandale, New Jersey, USA

THE initial derivation of atomic-scale models of the 4-connected framework crystal structures of zeolites and related materials is generally difficult. Interpretation of powder X-ray diffraction data to yield a unit cell and some symmetry information is often straightforward. Chemical analyses and sorption experiments indicate the number of framework tetrahedra, n_T , present in the unit cell (T represents the framework tetrahedral species, such as Si or Al) and the approximate pore dimensions. However, because synthetic zeolites are almost invariably microcrystalline, the possibilities for structure solution by conventional diffraction methods are limited, and framework structure determinations have traditionally relied heavily on the building of physical models. We describe here an alternative approach, in which approximate T-atom coordinates are derived from the unit-cell size and symmetry, and the value of n_T from computer modelling. An initially arbitrary T-atom configuration is optimized with respect to a 'cost function' based on the T-T distances, T-T-T angles and number of first-neighbour T-atoms, by simulated annealing using Monte Carlo methods. The potential of the method for structural determinations in both two dimensions (for structural projections) and three dimensions are illustrated by results obtained for a hexagonal cell of space group $P6/mmm$ and dimensions $a = 18.4 \text{ \AA}$ and $c = 7.5 \text{ \AA}$, and by determination of the previously unknown structure of lithium gallosilicate.

A large proportion of the ~65 known zeolite framework structure types¹ were solved based on an initial determination of the 4-connected T-atom framework by, for example, physical model building², by consideration of direct derivatives of known structures³, or by interpretation of electron microscopy and diffraction data^{4,5}. Oxygen atoms can be added at the midpoints of each of the T-T vectors, and the T- and O-atom coordinates then optimized readily by distance least-squares (DLS) based on established constraints on the T-O, O-O and T-T distances^{6,7}. The problem of structure solution reduces to one of defining approximate T-atom coordinates. The present approach describes the reasonableness of a given arrangement of the required number of T-atoms in the known unit cell in terms of a cost or 'energy' function and uses simulated annealing to adjust the T-atom coordinates to minimize that energy.

We define the total energy of a given T-atom configuration as the sum over the unique T-atoms of terms based on the T-T distances, T-T-T angles, the number, N_1 , of first neighbours, and (when T-atoms are possibly located on symmetry elements)

the number of symmetry operations that reproduce the original T-atom (Fig. 1). The forms of the energy as a function of the T-T distances and of the T-T-T angles are derived from histograms of observed data from a representative number of known zeolite structure types. The neighbour term is calculated from the number of T-atoms within a defined shell. The merging contribution is calculated from the distance between symmetry-related T-atoms found at less than a defined separation; merged atoms are excluded from the calculations of the other energy terms. Merging is permitted provided the total number of T-atoms per unit cell is $\geq n_T$. Appropriate relative weights of the four types of constraint were derived largely by trial and error, with the parameters being treated as tunable in complicated cases. Appropriate adjustment of these various terms (particularly the neighbour term) permits the same method to be applied to other coordination environments such as those that occur, for example, in mixed octahedral-tetrahedral framework structures.

Optimization of the coordinates of the unique T-atoms (the full unit-cell T-atom complement is generated by application of the space-group operators) to minimize the calculated energy sum is achieved by Monte Carlo methods using simulated annealing^{8,9} based on an initially random configuration. During annealing low-energy, 4-connected configurations generated are stored for further evaluation (typically, 2-20 are produced in each run). Simple representations (projections down the three principal crystallographic directions) of the full unit-cell contents for each of the saved models are examined visually, followed, where appropriate, by further molecular graphics analyses, or by addition of framework oxygen atoms and DLS optimization and/or simulation of the powder-X-ray diffraction pattern(s). A typical run for a three-dimensional case (see below) with $n_{\text{unique}} = 2$ unique T-sites, 24 symmetry operators and $n_T = 36$ consumes ~60 min CPU time on a VAX8350, or ~60 s on a Cray XMP/14 se.

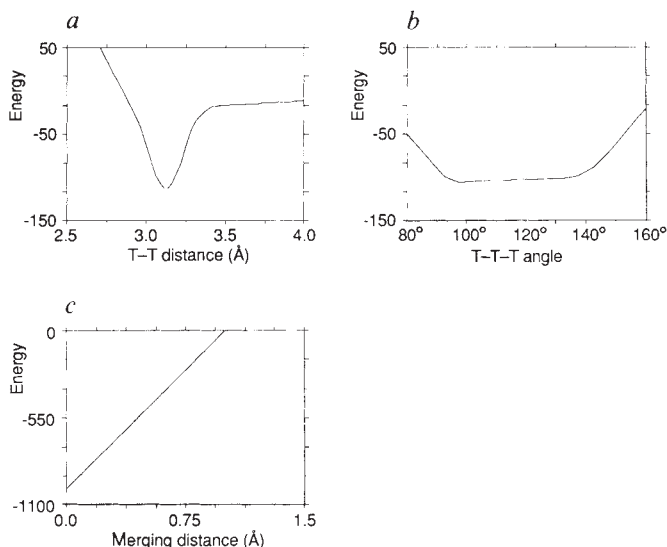


FIG. 1 Energy as a function of the T-T distance (a) and the T-T-T angle (b) used in the simulation procedure (calculated as smoothing spline fits to Boltzman equilibrium interpretations of histogrammed data taken from 32 representative zeolite crystal structures). Only the central portions are shown, the energy curves extrapolating linearly to energies of 1,000.0 and 0.0 at T-T distances of 0.0 and 5.0 Å, and to energies of 5,000.0 and 0.0 at T-T-T angles of 0.0 and 180° respectively. (c) The contribution to the energy sum for the merging of two symmetry-related atoms; merging is only permitted when the two atoms are at less than a defined minimum distance. The contribution to the energy sum from the coordination term is typically set to 1,000.0, 800.0, 600.0, 300.0, 0.0, 300.0, 600.0 for 0, 1, 2, 3, 4, 5 and 6 T-atoms within the defined first-neighbour radius (typically 3.7 Å) respectively.

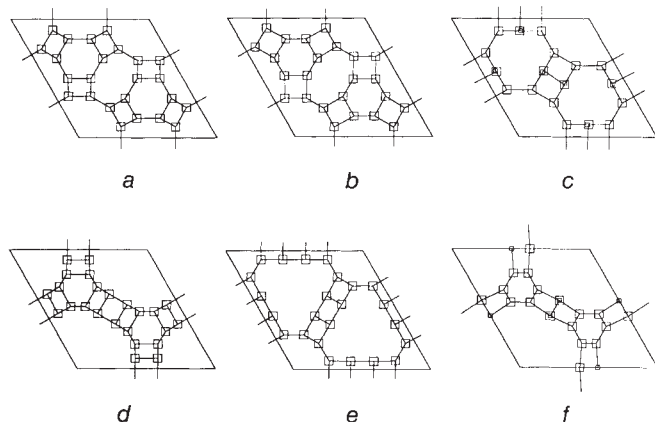


FIG. 2 The projection of the T-atom configuration in zeolite L (a) compared with two-dimensional modelling results (b-f) based on a hexagonal plane cell, $p6mm$, with $a=18.4$ Å (see text). Smaller squares indicate degenerate atoms.

In the known zeolite framework structure types¹, the main pores generally run parallel to a unit-cell axis, and in projection the T-atoms are reasonably dispersed and usually 3-connected. The present methods were, therefore, developed initially based on the (simpler) two-dimensional, projection case. The constraints on both T-T distances and T-T-T angles were relaxed substantially to enable proper treatment of T-T linkages oblique to the projection direction. Results for a plane cell with $p6mm$ symmetry (with 12 symmetry operators) and $a=18.4$ Å, for $n_{\text{unique}}=2$ and $n_T=24$, and for $n_{\text{unique}}=2$ and $n_T=18$, are shown in Fig. 2. A single run with $n_{\text{unique}}=2$ and $n_T=24$ yielded the *lhl* net (ref. 10) (Fig. 2b) expected for projections of the LTL and MAZ frameworks¹, and nets labelled *tsv* (Fig. 2d) and *twy* (Fig. 2e) by Smith¹⁰ (the projections of the three-dimensional

4-connected nets numbered 313 (ref. 11) or 81(2) (ref. 12), and 318 (ref. 11), respectively). The two-dimensional nets labelled *eo* (ref. 10) (Fig. 2f) (the projection of the 81(1) or 520 net (ref. 13) adopted by the 18-ring aluminophosphates VPI-5 (ref. 14) and $\text{ALPO}_4\text{-54}$ (ref. 13) and *tfn* (ref. 10) (Fig. 2c) were produced in a run based on $n_{\text{unique}}=2$ and $n_T=18$. Similarly successful results (in a series of possible hexagonal plane symmetries) were obtained for a smaller cell with $a=13.3$ Å that is typical of materials in the ABC-6 family of zeolites¹⁵. The projection approximation, however, does not converge successfully in cases such as the SOD-framework¹, where connected and non-connected vertices have, in projection, similar separations.

The full three-dimensional case is of greater utility but involves a larger number of independent variables. Annealing parameters that have been adjusted to facilitate convergence include the temperature decrement between successive cooling stages (generally chosen based on the evolution with temperature of the system specific heat⁸), and the temperature dependence of the maximum T-atom displacement in the Monte Carlo jiggle step. Typical results are illustrated in Fig. 3 for a hexagonal cell with $P6/mmm$ symmetry (24 symmetry operators), $a=18.4$ Å, $c=7.5$ Å, and with $n_{\text{unique}}=2$ and $n_T=36$ that is observed for zeolite L. A total of ten different runs generated, in aggregate but with much duplication, the LTL-framework¹ (Fig. 3a), the hypothetical framework initially proposed for zeolite omega¹⁶ (Fig. 3c), frameworks corresponding to nets numbered 313 (Fig. 3b), 318 (Fig. 3d) and 315 (Fig. 3e) described previously¹¹, a related framework with a circuit symbol $(4^36^3)_1(4^36^3)_2$ (Fig. 3f), and two sheet structures (Fig. 3g that has $n_T=48$, and Fig. 3h) that were observed with relatively high energies only at high simulation temperatures. These topologies correspond in each case to one of the two-dimensional nets deduced in the projection approximation (Figs 3a, 3c and 3g correspond to Fig. 2b; Fig. 3b to Fig. 2f, Figs 3d and 3f to Fig. 2d; Figs 3e and 3h to Fig. 2e). An example of an apparently more constrained system is provided by the high silica, 10-ring zeolite Theta-1 for which a single run (that consumed 15 min Cray XMP/14 se CPU time) based on the orthorhombic space group symmetry $Cmc2_1$, with $a=13.8$ Å, $b=17.4$ Å, $c=5.0$ Å, $n_{\text{unique}}=4$ and $n_T=24$ converged to a single 4-connected topology identical to that described for the TON-framework in 1984¹⁷.

Analysis of data for a condensed lithium gallosilicate that was recently described to be orthorhombic, space group $Pnam$ or $Pna2_1$, with $a=10.04$ Å, $b=6.63$ Å, $c=4.96$ Å (ref. 18) illustrates the solution of an unknown structure by the present method. Two runs based on this unit cell, space group $Pna2_1$, $n_{\text{unique}}=2$ and $n_T=8$, produced some 20 trial structures, which on inspection were found to represent different settings of three distinct 4-connected topologies. The correct model (Fig. 4) was identified by comparing the observed powder X-ray diffraction pattern with those simulated on the basis of the three models

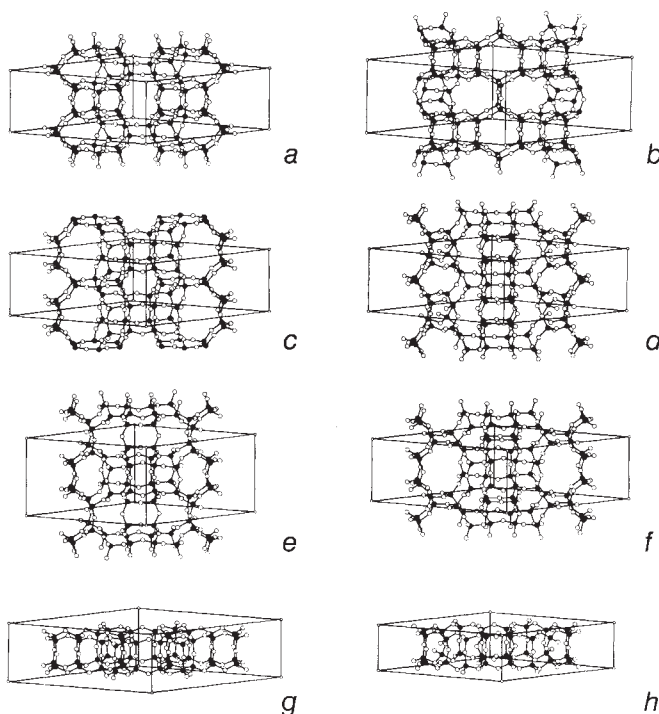


FIG. 3 Ball and stick representations of the eight distinct structure types produced in three-dimensions based on a hexagonal unit cell, $P6/mmm$, with $a=18.4$ Å, $c=7.5$ Å, $n_{\text{unique}}=2$ and $n_T=36$ that is observed for zeolite L (see text). The structures are shown following addition of the framework oxygen atoms and optimization of the atomic coordinates and the unit-cell dimensions by conventional distance least squares^{6,7} (T-atoms, solid; O-atoms, open).

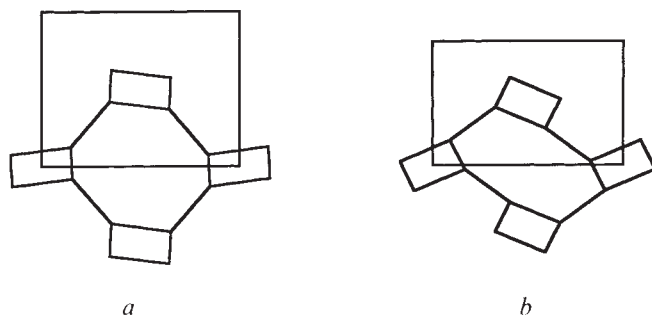


FIG. 4 The [001] projections of the framework structures of gallosilicate zeolite Li-A(BW) (a) compared with the DLS-optimized framework structure of its calcination product (b) solved by the present method.

with oxygen atoms entered (approximately) at the midpoints of the T-T vectors. This structure proves to be isotopological with gallosilicate zeolite Li-A(BW) of which it is a calcination product¹⁸. Removal of the sorbed water results in gross readjustment of the 8-ring channel, a 20% reduction in the length of the *b*-axis (Fig. 4), and a powder X-ray diffraction pattern that bears no resemblance to that of the parent. Although no sorption capacity for water was noted in this material¹⁸, reversion to the parent hydrated structure may occur over extended periods under suitable water vapour pressures.

The present algorithms apparently work well for small numbers of unique T-atoms, $n_{\text{unique}} < \sim 6$ (which encompasses >85% of the known zeolite structure types in their highest possible

symmetries¹), where a small number of passes through the program generally yields the sought solution(s). For larger numbers of independent T-atoms, especially when merging is required, many 4-connected topologies are produced. The present energy function is apparently not yet sufficiently restrictive in these more complicated cases. By evaluating results for a larger number of known materials, however, it is hoped to further improve the constraint functions and the parameters controlling passage through the simulated annealing process. Although certainly not guaranteed to work in all cases, these procedures provide a new approach to the difficult process of zeolite structure solution, and a rapid means of extending studies of hypothetical 4-connected nets. □

Received 5 July; accepted 3 October 1989.

- Meier, W. M. & Olson, D. H. *Atlas of Zeolite Structure Types* (Butterworths, Surrey, 1987).
- LaPierre, R. B. *et al. Zeolites* **5**, 346–348 (1985).
- Bennett, J. M., Richardson, J. W., Pluth, J. J. & Smith, J. V. *Zeolites* **7**, 160–162 (1987).
- Treacy, M. M. J. & Newsam, J. M. *Nature* **332**, 249–251 (1988).
- Newsam, J. M., Treacy, M. M. J., Koetsier, W. T. & deGruyter, C. B. *Proc. R. Soc. Lond. A* **420**, 375–405 (1988).
- Meier, W. M. & Villiger, H. Z. *Kristallogr.* **129**, 411–423 (1969).
- Baerlocher, Ch., Hepp, A. & Meier, W. M. *DLS-76 Manual* (ETH, Zurich, 1977).
- K. Binder (ed.) *Applications of the Monte Carlo Method in Statistical Physics* 2nd edn (Springer, Berlin, 1987).
- Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. *Science* **220**, 671–680 (1983).

- Smith, J. V. in *Zeolites: Facts, Figures, Future* (eds van Santen, R. A. & Jacobs, P. A.) 29–47 (Elsevier, Amsterdam, 1989).
- Smith, J. V. & Dytrych, W. J. Z. *Kristallogr.* **175**, 31–36 (1986).
- Smith, J. V. & Dytrych, W. J. *Nature* **309**, 607–608 (1984).
- Richardson, J. W., Smith, J. V. & Pluth, J. J. *J. phys. Chem.* (in the press).
- Davis, M. E., Saldarriaga, C., Montes, C., Garces, J. & Crowder, C. *Nature* **331**, 698–699 (1988).
- Smith, J. V. & Bennett, J. M. *Am. Miner.* **66**, 777–788 (1981).
- Barrer, R. M. & Villiger, H. *J. chem. Soc. chem. Commun.*, 659–660 (1969).
- Barri, S. A. I., Smith, G. W., White, D. & Young, D. *Nature* **312**, 533–534 (1984).
- Newsam, J. M. *J. phys. Chem.* **92**, 445–452 (1987).

ACKNOWLEDGEMENTS. We thank J. R. Chelikowsky, M. M. J. Treacy and A. J. Jacobson for discussions.

Strontium, neodymium and lead isotope constraints on near-ridge seamount production beneath the South Atlantic

Pat Castillo*[‡] & Rodey Batiza†

* Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch, Washington, DC 20015, USA

† Hawaii Institute of Geophysics, University of Hawaii, Honolulu, Hawaii 96822, USA

STUDIES^{1–7} of seamounts near the East Pacific Rise (EPR) have shown that, although most seamount lavas are petrographically and chemically identical to mid-ocean-ridge basalt, they are chemically and isotopically more diverse than those erupted on the rise axis. They are also generally more primitive (higher MgO content) and, in some cases, more depleted in incompatible elements than the axial basalts. This indicates that although near-ridge seamounts and the EPR share a common mantle source, there must be significant differences in their magma-supply processes. Seamounts are probably built by small batches of melt that rise rapidly to the surface, preserving evidence of heterogeneity in the mantle source region. By contrast, the processes of melting and melt segregation, storage and extrusion along the fast-spreading (9–13 cm yr^{−1}) EPR are more efficient in mixing and homogenizing basalt compositions^{1–7}. Here we show that the relationship of strontium, neodymium and lead isotope data between seamounts in the South Atlantic and the nearby axis of the slow-spreading Mid-Atlantic Ridge (MAR) is similar to that seen in the Pacific. This indicates that the processes leading to formation of near-ridge seamounts are similar at a wide range of spreading rates. Differences in the specific isotope signatures of lavas from near-ridge seamounts and axes of the EPR and MAR reflect regional differences in the upper-mantle source of mid-ocean-ridge basalts.

The new Sr, Nd and Pb isotope data presented here (Table 1) are mostly from glassy basalts that were dredged from seamounts near the MAR during cruise RC 2802 (RV *Conrad*) in 1987^{8,9} (Fig. 1). For comparison, we also present a new graphical representation of isotope data for fresh glassy basalts dredged at an average spacing of ~7 km along the MAR from the Rio Grande (25°40' S) to the Moore (26°30' S) fracture zones. This segment of the MAR at 26° S is spreading slowly (3–4 cm yr^{−1}) and is far from active hotspot volcanism. Indeed, all lavas from the MAR at 26° S and near-ridge seamounts are 'normal' mid-ocean-ridge basalts (MORBs) in terms of major and incompatible element contents^{8,9}. As in the EPR, seamount basalts are systematically richer in MgO and more depleted in TiO₂ contents than most of the axial lavas. Major-element data indicate that the seamount basalts and some high-MgO axial lavas from near the fracture zones and the axial high (Fig. 1) were produced by smaller extents of partial melting at slightly higher pressure than most of the axial lavas^{8,9}.

Seamounts basalts have more variable ⁸⁷Sr/⁸⁶Sr and ¹⁴³Nd/¹⁴⁴Nd values than the axial lavas in the South Atlantic (Table 1 and Fig. 2a). Moreover, Sr isotopic ratios are generally higher and Nd isotopic ratios are lower for seamount basalts than for axial lavas. For example, samples 9 and 10, which are from one seamount, range from 0.70260 to 0.70350 in ⁸⁷Sr/⁸⁶Sr and from 0.51297 to 0.51314 in ¹⁴³Nd/¹⁴⁴Nd values. Sr and Nd isotopic ratios of samples 5, 6 and 7, also from a single seamount and related cones, are less variable but these ratios still are more diverse than those of axial lavas from the MAR nearby that vary from only 0.70247 to 0.70259 in ⁸⁷Sr/⁸⁶Sr and from 0.51308 to 0.51314 in ¹⁴³Nd/¹⁴⁴Nd values (P. C., unpublished data).

Except for sample 9, ²⁰⁶Pb/²⁰⁴Pb values of seamount basalts (18.21–18.33) are well within the range of the axial lavas (18.16–18.40) (Table 1 and Fig. 2b). In turn, the Pb isotope signature of the axial lavas is very similar to that of other normal-MORBs from the 22° S to 31° S segment of the MAR¹⁰. Seamount basalts, however, have distinctly more variable and systematically higher ²⁰⁸Pb/²⁰⁴Pb and, to a lesser extent, ²⁰⁷Pb/²⁰⁴Pb, for given ²⁰⁶Pb/²⁰⁴Pb values than the axial lavas (37.86–38.17 against 37.73–38.07, and 15.47–15.55 against 15.48–15.53, respectively). Sample 9 has high ²⁰⁷Pb/²⁰⁴Pb (15.66) but approximately normal ²⁰⁸Pb/²⁰⁴Pb (38.48) for a given ²⁰⁶Pb/²⁰⁴Pb value relative to axial lavas (Fig. 2b).

[‡] Present address: Rosenstiel School of Marine and Atmospheric Science, University of Miami, 4600 Rickenbacker Causeway, Miami, Florida 33149, USA.

TABLE 1 Sr, Nd, and Pb isotopic ratios and Sr and Nd contents of representative seamount and axial lavas

Sample	$^{87}\text{Sr}/^{86}\text{Sr}$	$^{143}\text{Nd}/^{144}\text{Nd}$	$^{206}\text{Pb}/^{204}\text{Pb}$	$^{207}\text{Pb}/^{204}\text{Pb}$	$^{208}\text{Pb}/^{204}\text{Pb}$
1-1	0.702595	0.513093 (9)	18.244 (21)	15.471 (21)	37.860 (49)
1-2wr	0.702553	0.513105 (22)			
5-5	0.703629	0.512919 (10)	18.269 (31)	15.548 (27)	38.173 (68)
6-3	0.703098	0.512973 (9)	18.334 (19)	15.523 (18)	38.152 (44)
6-4	0.703155	0.512954 (9)	18.269 (12)	15.535 (11)	38.143 (24)
7-4	0.703430	0.512988 (16)			
7-5	0.703475	0.512985 (13)	18.253 (12)	15.515 (11)	38.092 (24)
9	0.703496	0.512972 (9)	18.953 (0)	15.659 (0)	38.475 (1)
9dup			18.974 (1)	15.686 (1)	38.567 (2)
10-6wr	0.702646	0.513135 (17)			
10-7wr	0.702604	0.513107 (21)	18.208 (6)	15.483 (5)	37.826 (7)
12-29	0.702552	0.513086 (23)	18.304 (3)	15.512 (2)	37.960 (7)
15-2	0.702471	0.513129 (14)	18.126 (2)	15.482 (2)	37.731 (3)
18-3	0.702525	0.513120 (17)	18.397 (2)	15.531 (2)	38.068 (7)
20-1	0.702538	0.513087 (19)	18.262 (3)	15.483 (2)	37.858 (6)

Sr isotopic ratios are fractionation corrected to $^{86}\text{Sr}/^{88}\text{Sr}=0.1194$ and normalized to $^{87}\text{Sr}/^{86}\text{Sr}=0.710265$ for NBS 987; Nd isotopic ratios are fractionation corrected to $^{146}\text{Nd}/^{144}\text{Nd}=0.72225$ ($^{146}\text{Nd}/^{144}\text{Nd}=0.7219$) and are reported relative to $^{143}\text{Nd}/^{144}\text{Nd}=0.511860$ for the La Jolla Standard. Isotopic composition of Pb are corrected for mass fractionation based on average measured values for NBS 981 using the values given by Catanzaro *et al.*³⁰. All analyses are done on fresh glasses except for 3 samples (indicated by wr) which are done on leached whole-rock powders. Sample 9dup is an analysis on a different glass piece of sample 9. Numbers in parentheses are in-run analytical precisions and refer to the last significant digits; in-run precision for Sr isotope measurements < 10.

These results indicate that despite large differences in spreading rate, both the Atlantic and Pacific display similar axis-seamount isotope systematics; seamounts are more diverse isotopically than axial lavas. As the abundances, morphology, and major-element contents of MAR near-axis seamounts are also similar to their Pacific counterparts^{8,9}, it seems likely that near-axis seamounts are produced by similar processes operating at a wide range of spreading rates. Because of the gross petrological similarity of lavas from seamounts and nearby axis, the close proximity of seamounts to the axis, and because the zone of mantle upwelling is thought to be ~100 km wide^{11,12}, it is very clear that axial and seamount melts tap the same mantle source. However, the systematic difference between axial and seamount lavas in major elements, trace elements and isotopes indicates that this common source is tapped in different ways and/or the melt migration and/or the storage history of seamount magmas and axial magmas are different.

As already mentioned, the petrological diversity but generally high MgO contents of near-ridge seamount lavas indicate that seamounts tap relatively small volumes of mantle, which rise quickly without extensive modifications¹⁻⁷. In addition, seamount magmas may be the result of smaller extents of partial melting, in which case magmas may contain high proportions of more enriched, more easily melted components of the upper mantle^{1,5,12} than axial lavas. By contrast, axial magmas may be mixed and homogenized during melt segregation, melt migration or storage in crustal magma chambers.

Despite the similarity in axis-seamount systematics in the Atlantic and Pacific there are also significant differences. For example, Atlantic near-ridge seamounts show much more variability in $^{87}\text{Sr}/^{86}\text{Sr}$ than do the Pacific ones (Fig. 2a). Atlantic seamounts exhibit smaller ranges of $^{143}\text{Nd}/^{144}\text{Nd}$ and $^{206}\text{Pb}/^{204}\text{Pb}$ variation than Pacific ones, but this could be due to the small data base for Atlantic near-ridge seamounts. Note also that Atlantic seamount samples are all normal MORBs whereas in the Pacific, most of the high $^{87}\text{Sr}/^{86}\text{Sr}$ and low $^{143}\text{Nd}/^{144}\text{Nd}$ samples are alkali basalts and their differentiates⁷. The reason for this is not entirely clear, but it could be due to differences in the depth of origin or age of eruption of Atlantic and Pacific near-axis seamounts. Alternatively, the geochemical characteristics beneath the South Atlantic and the Pacific might be different¹³.

Another notable feature of the data is that except for some lavas from the Lamont seamount chain in the Pacific⁶, Atlantic near-axis seamounts have lower mean $^{206}\text{Pb}/^{204}\text{Pb}$ than Pacific ones (Fig. 2b). The Lamont samples, however, have low

$^{87}\text{Sr}/^{86}\text{Sr}$, whereas the low $^{206}\text{Pb}/^{204}\text{Pb}$ Atlantic samples have high $^{87}\text{Sr}/^{86}\text{Sr}$ (Fig. 3b). The overall pattern of lower $^{206}\text{Pb}/^{204}\text{Pb}$ and lower $^{143}\text{Nd}/^{144}\text{Nd}$, but higher $^{207}\text{Pb}/^{204}\text{Pb}$ and $^{208}\text{Pb}/^{204}\text{Pb}$ for a given $^{206}\text{Pb}/^{204}\text{Pb}$ of Atlantic seamounts relative to Pacific ones (Figs 2b, 3a) is apparently paralleled by a similar isotope pattern at the respective ridge axes. The MAR in the South Atlantic has a lower $^{206}\text{Pb}/^{204}\text{Pb}$ than the EPR (Fig. 2b). Like-

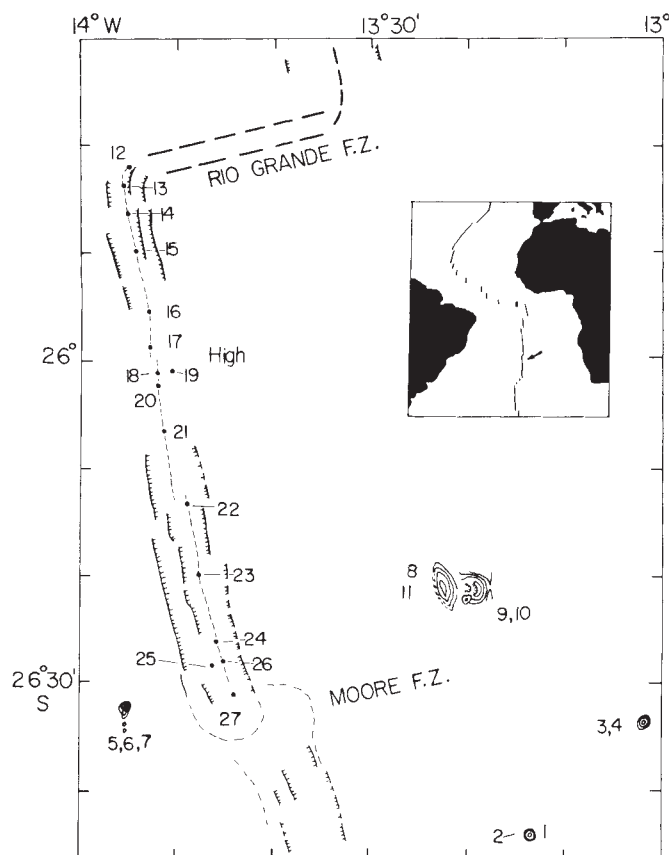


FIG. 1 Schematic map of the Mid-Atlantic Ridge between the Rio-Grande and Moore fracture zones, showing dredges on near-ridge seamounts (1-11), along the MAR axis (12-18, 20-24, 26, 27) and slightly off-axis (19, 25) (from ref. 8). The ridge axis is elevated near the centre of the segment at ~26°S as indicated.

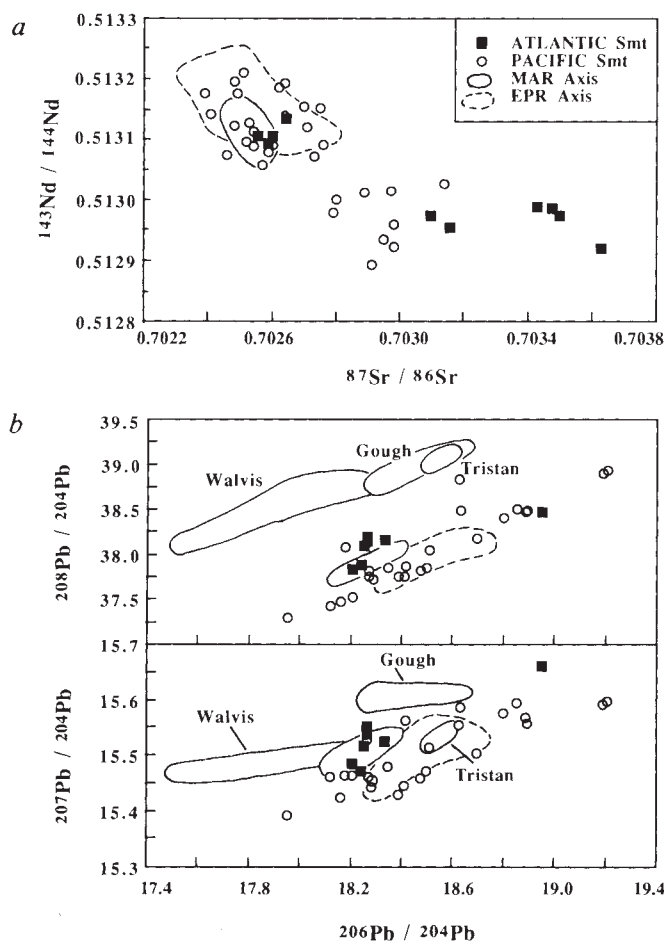


FIG. 2 a, Plot of $^{143}\text{Nd}/^{144}\text{Nd}$ against $^{87}\text{Sr}/^{86}\text{Sr}$ for lavas from near-ridge seamounts in the Atlantic and Pacific. Isotope data for axial lavas shown for reference are restricted to those segments of the MAR and EPR that are adjacent to the seamounts. These isotope data are from the MAR between $25^{\circ}40'$ and $26^{\circ}30'$ S (representative samples from dredges 12–27 in Fig. 1) and from the EPR between 8° to 22° N latitude. b, plots of $^{208}\text{Pb}/^{204}\text{Pb}$ and $^{207}\text{Pb}/^{204}\text{Pb}$ against $^{206}\text{Pb}/^{204}\text{Pb}$ for the samples shown in a. Symbols and reference fields as in a. (Pb isotope data for an axial lava from the MAR at $26^{\circ}00'$ S reported by Hanan and Schilling¹⁰ (EN063 12D-5g) plot inside the field for MAR axis.) Note that although almost all lavas from near-ridge seamounts and the ridge axis in the south Atlantic have similar $^{206}\text{Pb}/^{204}\text{Pb}$, seamount basalts have higher $^{208}\text{Pb}/^{204}\text{Pb}$ and $^{207}\text{Pb}/^{204}\text{Pb}$ than axial lavas. This Pb isotopic signature is a common feature of lavas from the Tristan da Cunha and Gough hot spots and their apparent trace, the Walvis Ridge, further to the south. Data from near-ridge seamounts in the Pacific with lowest $^{206}\text{Pb}/^{204}\text{Pb}$ values are from the Lamont group of seamounts⁶. Despite the fewer number of available samples from the South Atlantic, near-ridge seamounts clearly are more isotopically heterogeneous than axial lavas.

wise, the MAR in the South Atlantic has a lower mean $^{143}\text{Nd}/^{144}\text{Nd}$ but higher mean $^{207}\text{Pb}/^{204}\text{Pb}$ and $^{208}\text{Pb}/^{204}\text{Pb}$ than the EPR (Fig. 2b). These observations are consistent with the notion of a similar mantle source for both axial and seamount magmas.

With the exception of some samples from near the Oceanographer Fracture Zone¹⁴ and near the Equator¹⁵, the MAR in the North Atlantic exhibits an isotope pattern different from that in the South Atlantic^{10,13}. For example, North Atlantic axial MORBs far from hotspots generally have similar $^{206}\text{Pb}/^{204}\text{Pb}$ and higher $^{143}\text{Nd}/^{144}\text{Nd}$ than the EPR axial lavas^{16,17}. These apparent differences between the MAR in the North and South Atlantic may be due to regional differences in the mantle characteristics¹³, but a possible plume contribution in the South Atlantic¹⁰ is difficult to rule out.

The comparison between the South MAR and EPR indicates

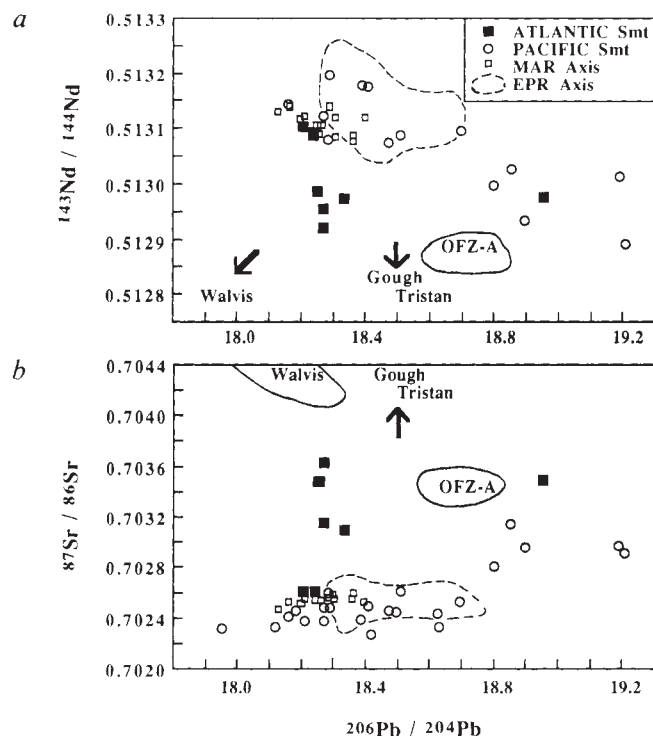


FIG. 3 Plots of $^{143}\text{Nd}/^{144}\text{Nd}$ against $^{206}\text{Pb}/^{204}\text{Pb}$ (a) and $^{87}\text{Sr}/^{86}\text{Sr}$ against $^{206}\text{Pb}/^{204}\text{Pb}$ (b) for lavas from near-ridge seamounts in the Atlantic and Pacific. Axial lavas from the MAR at 26° S generally have lower mean values of $^{206}\text{Pb}/^{204}\text{Pb}$ and lower mean values of $^{143}\text{Nd}/^{144}\text{Nd}$ but higher mean values of $^{207}\text{Pb}/^{204}\text{Pb}$ and $^{208}\text{Pb}/^{204}\text{Pb}$ for given $^{206}\text{Pb}/^{204}\text{Pb}$ than axial lavas from the EPR between 8° – 22° N. These regional isotope differences generally are more pronounced in the isotopic signature of lavas from near-ridge seamounts. Nd isotopic ratios of the Lamont seamount basalts are not available; nevertheless, the fairly good negative correlation between Nd and Sr isotopic ratios of other seamounts near the EPR^{5,7} indicate that these lavas also have lower $^{143}\text{Nd}/^{144}\text{Nd}$ for given $^{206}\text{Pb}/^{204}\text{Pb}$ ratios than Atlantic seamount and axial lavas. Fields for the Tristan da Cunha, Gough, Walvis Ridge, and 'type A' Oceanographer Fracture Zone lavas¹³ are shown for reference. Note that all near-ridge seamount and axial isotope data can be enclosed by a single large triangle (not shown) in these two diagrams.

that near-axis seamount lavas tend to magnify isotope differences that are present but more subdued in axial lavas. In the southern MAR, our new data confirm the presence of at least three mantle components as proposed by Hanan *et al.*¹⁰. Two of these are also present at the EPR and the third is a component enriched in the Tristan–Gough–Discovery hotspots (Fig. 3a,b). The presence of this component in the near-ridge seamount and in the MAR axial lavas is difficult to explain. The large geographic distance between the MAR at 26° S and the hotspots (31° – 47° S) is inconsistent with present models of hotspot–ridge interaction^{10,18,19}. Moreover, this component is also recognized in MORBs from the North Atlantic^{14,15} and elsewhere in the Indian Ocean^{20,21} posing further difficulty for a simple hotspot contamination model. Our preferred hypothesis^{13,22} is that the component is pervasive in areas underlain by low-velocity, presumably hot, regions in the mantle²³ recognized by tomographic studies²⁴. These regions also are correlated with hotspot concentration and geoid anomalies^{25,26}. One such region is below the southern tip of Africa but extends north to the Azores^{13,23}.

New isotope data also add further evidence that the suggested inverse relationship between spreading rate and isotope variability^{27,28} is not always valid. Our results as well as those of others^{16,17,29} show that apparently there is no such relationship. Rather, both the fast-spreading EPR and the slow-spreading MAR are isotopically variable; the observed isotope difference is most probably due to regional differences of heterogeneous mantle components beneath spreading centres. □

1. Batiza, R. & Vanko, D. *J. geophys. Res.* **26**, 564–602 (1984).
2. Allan, J. F., Batiza, R. & Lonsdale, P. *Seamounts, Islands, and Atolls*, *Geophys. Monogr.* No. 43 (eds Keating, B. H. et al.) 255–282 (Am. geophys. Un., Washington, DC, 1987).
3. Allan, J. F., Batiza, R., Perfit, M. R., Fornari, D. J. & Sack, R. O. *J. Petrol.* (submitted).
4. Fornari, D. J., Perfit, M. R., Allan, J. F. & Batiza, R. *Nature* **331**, 511–513 (1988).
5. Zindler, A., Staudigel, H. & Batiza, R. *Earth planet. Sci. Lett.* **70**, 175–195 (1984).
6. Fornari, D. J. et al. *Earth planet. Sci. Lett.* **89**, 63–83 (1988).
7. Graham, D. W. et al. *Contr. Miner. Petr.* **99**, 446–463 (1988).
8. Batiza, R., Melson, W. G. & O'Hearn, T. *Nature* **335**, 428–431 (1988).
9. Batiza, R. et al. *J. Geol.* **97**, 209–220 (1989).
10. Hanan, B. B., Kingsley, R. H. & Schilling, J.-G. *Nature* **322**, 137–144 (1986).
11. McKenzie, D. P. *Earth planet. Sci. Lett.* **74**, 81–91 (1985).
12. Sleep, N. H. *J. geophys. Res.* **89**, 10029–10041 (1984).
13. Castillo, P. *Nature* **336**, 667–670 (1988).
14. Shirey, S., Bender, J. F. & Langmuir, C. H. *Nature* **325**, 217–223 (1987).
15. Schilling, J.-G., Hanan, B. B. & McCully, B. *Eos* **69**, 1426 (1988).
16. White, W. M., Hofmann, A. W. & Puchelt, H. *J. geophys. Res.* **92**, 4881–4893 (1987).
17. Ito, E., White, W. M. & Gopel, C. *Chem. Geol.* **62**, 157–176 (1987).
18. Schilling, J.-G. *Earth planet. Sci. Lett.* **25**, 103–115 (1975).
19. Schilling, J.-G. *Nature* **314**, 62–67 (1985).
20. Klein, E. M., Langmuir, C., Zindler, A., Staudigel, H. & Hamelin, B. *Nature* **333**, 623–629 (1988).
21. Mahoney, J. J. et al. *J. geophys. Res.* **94**, 4033–4052 (1989).
22. Castillo, P. *Eos* **69**, 490 (1988).
23. Silver, P. G., Carlson, R. W. & Olson, P. A. *Rev. Earth planet. Sci.* **16**, 477–541 (1988).
24. Dziewonski, A. M. *J. geophys. Res.* **89**, 5929–5952 (1984).
25. Hager, B. H. *J. geophys. Res.* **89**, 6003–6015 (1984).
26. Cazenave, A., Souriau, A. & Dominh, K. *Nature* **340**, 54–57 (1989).
27. Cohen, R. S. & O'Nions, R. K. *J. Petrol.* **23**, 299–239 (1982).
28. Allègre, C. J., Hamelin, B. & Dupre, B. *Earth planet. Sci. Lett.* **71**, 71–84 (1984).
29. Macdougall, J. D. & Lugmair, G. W. *Earth planet. Sci. Lett.* **77**, 273–284 (1986).
30. Catanzaro, E. J., Murphy, T. J., Shields, W. R. & Garner, E. L. *J. Res. natn. Bur. Stand.* **72A** (1968).

ACKNOWLEDGEMENTS. We thank the captain, crew, and scientific party (P. J. Fox and P. R. Vogt, Co-chief scientists) of RV *Conrad* for help during cruise RC 2802 and W. M. White, S. Shirey and D. Fisher for reviews. Support from the CIW to P.C. and NSF and ONR to R.B. is also acknowledged.

Young high-temperature granulites from the base of the crust in central Mexico

J. L. Hayob*, E. J. Essene*, J. Ruiz†, F. Ortega-Gutiérrez‡ & J. J. Aranda-Gómez§

* Department of Geological Sciences, University of Michigan, Ann Arbor, Michigan 48109, USA

† Department of Geosciences, University of Arizona, Tucson, Arizona 85721, USA

‡ Instituto de Geología, Universidad Nacional Autónoma de México, México City, México

§ Instituto de Geología, Universidad Nacional Autónoma de México, Guanajuato, México

CHARACTERIZATION of the lower continental crust is essential to understanding processes of crustal formation and evolution. Many workers have concluded that the lower crust is composed of granulite-facies rocks (see, for example, refs 1, 2), but many exposed granulite-facies terranes seem to have mid-crustal rather than lower-crustal origins^{3–5}. Here we present new data from direct samples of the lower continental crust, in the form of metapelitic rock fragments brought to the surface by Quaternary volcanism in central Mexico. These xenoliths contain exsolved ternary feldspars that record extremely high minimum metamorphic temperatures of 950–1,100 °C, the highest yet known to be preserved in deep-seated metamorphic rocks. The metamorphism seems to be the result of underplating by basaltic magma and to be regional in extent. Our barometry on these samples demonstrates that granulite-facies metamorphic rocks are present in the lower crust, and provides important constraints on tectonic models and interpretations of geophysical data. Chemical and textural evidence suggest that the xenoliths were not able to cool to below 850–900 °C before being entrained in their host magmas, indicating that the granulite-facies metamorphism occurred no more than ~30 Myr ago.

So far most studies of xenoliths have focused on characterization of samples from the Earth's upper mantle⁶, and only limited mineralogical and geochemical data have been obtained

on xenoliths believed to be from the lower crust. In addition, most studies of deep crustal xenoliths have concentrated on mafic and intermediate orthogneisses⁶. Although modally less abundant than mafic orthogneisses, xenoliths with sedimentary protoliths may be more useful in providing definitive pressure information, because their mineralogy is often more appropriate for applying existing geobarometers^{7–10}.

Xenoliths of garnet-sillimanite paragneiss, occurring in Quaternary basaltic tuffs¹¹ from the central Mexican plateau in the state of San Luis Potosí, contain the assemblage garnet-sillimanite-mesoperthite-quartz-rutile-graphite, ±plagioclase-ilmenite-zircon-monazite. The paragneisses are the major lithology at El Toro and represent ~10% of the xenolith population at La Joya Honda and Los Palau; mafic and felsic orthogneisses are the dominant lithologies at these two localities. Here we present the pressure (*P*)-temperature (*T*) conditions of last equilibration for the paragneisses, and discuss the crustal level from which these samples were derived and the information they reveal about the crust beneath central Mexico.

The paragneiss xenoliths were selected for detailed study based on mineralogical and textural similarities and the presence of univariant mineral assemblages with respect to barometric reactions. Five samples are from the El Toro (ET) cinder cone, three are from La Joya Honda (LJH) maar, and two are from Los Palau (LP) maar. Field relations are given elsewhere^{11–13}. The xenoliths range from 4 to 10 cm in diameter and appear fresh. In general, all samples have similar mineral chemistries. None of the selected samples exhibit distinct layering or mineral zoning. Evidence of partial fusion is given by the ubiquitous occurrence of devitrified melt rims around garnets (see, for example, refs 14, 15). Quench textures in the melt rims consist of intergrowths of euhedral orthopyroxene, hercynite-spinel_{ss}, and interstitial glass and/or calcic plagioclase. Stähle¹⁴ has suggested that the breakdown of almandine garnet to orthopyroxene + hercynite + glass occurs rapidly at ~1,100 °C and 0.1–0.3 GPa.

The compositions of the feldspars may provide constraints on the thermal histories of the xenoliths. Plagioclase lamellae occurring within an alkali feldspar host and large plagioclase crystals not hosted by an alkali feldspar have very similar anorthite contents, within a given sample. The plagioclase lamellae are typically slightly more potassic, indicating that the larger plagioclase crystals formed during cooling from an initially homogeneous alkali feldspar¹⁶. Exsolution lamellae in feldspars are typically 10–30 µm in width. Exsolved feldspar compositions are given in Fig. 1a and Tables 1 and 2. The calculated

TABLE 1 Representative host, lamellae and reintegrated feldspar analyses from a single mesoperthite crystal in sample ET24

	Kfs host	PI lam.	Reintegrated		
SiO ₂	64.15	58.98	60.64	60.85	61.22
Al ₂ O ₃	19.51	25.53	22.98	23.33	23.47
Fe ₂ O ₃	0.01	0.19	ND*	ND*	ND*
CaO	0.77	7.41	4.30	4.42	4.81
Na ₂ O	3.12	6.42	5.21	5.24	5.30
K ₂ O	11.60	1.56	5.82	5.45	5.23
Total	99.16	100.09	98.95	99.29	100.03

Formulae normalized to 5 cations

Si	2.94	2.63	2.76	2.76	2.76
Al	1.06	1.35	1.23	1.25	1.25
Fe ³⁺	0.00	0.01	ND*	ND*	ND*
Ca	0.04	0.36	0.21	0.21	0.23
Na	0.28	0.56	0.46	0.46	0.46
K	0.68	0.09	0.34	0.32	0.30
O	7.99	7.98	7.97	7.99	8.00

* Fe₂O₃ was not determined (ND) for the reintegrated analyses. Kfs, alkali-feldspar, PI lam, plagioclase lamella.

TABLE 2 Component mole percentages for each sample from averaged mineral analyses

Sample number	Garnet			Rutile TiO ₂	Ilmenite FeTiO ₃	Fs host		Fs lam.		Integ.		Temperature (°C)	
	Al	Py	Gr			An	Or	An	Or	An	Or	Int.	Exs.
ET11	55	39	5	98	95	4	73	39	10	20	40	1,100	880
ET16	52	43	4	98	—	3	69	34	10	10	54	1,000	890
										23*	30*	1,050	890
ET24	49	46	4	98	98	4	69	35	9	22	32	1,050	880
ET38	55	40	4	97	—	3	73	36	10	23	33	1,075	880
ET42	55	40	4	97	—	4	65	30	12	18	34	1,025	890
LJH14	49	45	4	97	—	3	66	32	14	15	44	1,050	910
LJH15	54	42	3	98	—	6	52	27	17	19	28	1,000	965
LJH32	53	43	3	98	89	6	51	26	15	19	26	975	940
LP7	54	42	3	98	84	5	59	33	14	23	30	1,050	950
LP8	50	45	3	97	81	4	56	23	24	9	49	950	950

Sample locality abbreviations are given in the text.
Mineral abbreviations: Al, almandine; Py, pyrope; Gr, grossular; An, anorthite; Or, orthoclase; Fs, feldspar; Lam., lamellae. Spessartine and albite can be calculated by difference from 100. Int., average minimum reintegrated temperature, and Exs., average temperature of exsolution using the model of Lindsley and Nekvasil¹⁷.
* Bimodal distribution of reintegrated analyses observed for this sample.

tie lines for 900 °C from the ternary feldspar model of Lindsley and Nekvasil¹⁷ are shown at *P* = 1.0 GPa; the pressure dependence of the feldspar equilibria is from ref. 18. The mesoperthites contain lamellar plagioclase (An₂₀₋₄₀Ab₄₉₋₆₂Or₈₋₂₈) and alkali feldspar (An₃₋₇Ab₂₀₋₄₅Or₄₇₋₇₆) (Tables 1 and 2). Graphical use of the ternary two-feldspar thermometer in ref. 17 yields temperatures of ~800–950 °C for lamellae and host feldspars (Fig. 1*a* and Table 2). These temperatures are much higher than any recorded by exsolved feldspars of exposed granulites^{16,19}. Reintegration of the mesoperthites by image analysis and broad-beam energy-dispersive analysis yields former ternary feldspar compositions of An₆₋₂₄Ab₃₅₋₅₇Or₂₂₋₅₅ (Tables 1 and 2). The compositions are among the most ternary reported from rocks of undoubted metamorphic origin^{16,20-23}. The isotherms of the ref. 17 model require minimum temperatures of 950–1,125 °C (Table 2 and Fig. 1*b*) for the reintegrated feldspars to be stable.

Tie lines connecting exsolved alkali feldspar and plagioclase compositions for El Toro samples agree well with calculated tie lines¹⁷ (Fig. 1*a*). La Joya Honda and Los Palau samples give mixed results for exsolved compositions, as tie lines cross, indicating that these compositions may not represent equilibrium assemblages. Nevertheless, the reintegrated values give similar minimum temperatures of 950–1,125 °C for the three localities. Regardless of the details of exsolution processes, these mesoperthites yield minimum reintegrated temperatures that are considerably higher than those typically obtained for exposed granulites^{16,19}.

Figure 2 is a back-scattered electron image of a mesoperthite from sample ET24; quartz and rutile are also present. The regular and continuous exsolution texture is typical of the mesoperthites, and is interpreted as evidence for slow cooling after regional metamorphism. If the mesoperthites were produced during or after transport to the surface it is highly unlikely that the intergrowths would be so coarse and regular. The homogeneity of lamellae and host compositions within each xenolith suggests that equilibrium between the feldspars was commonly attained, and no chemical gradients across lamellae have been identified with the electron microprobe. The occurrence of mesoperthite inclusions in unzoned garnets from three samples (LP7, LP8, LJH32) seems to exclude an origin by epitaxial growth, as proposed²³ for mesoperthites from Kilbourne Hole (New Mexico), and indicates that the garnets

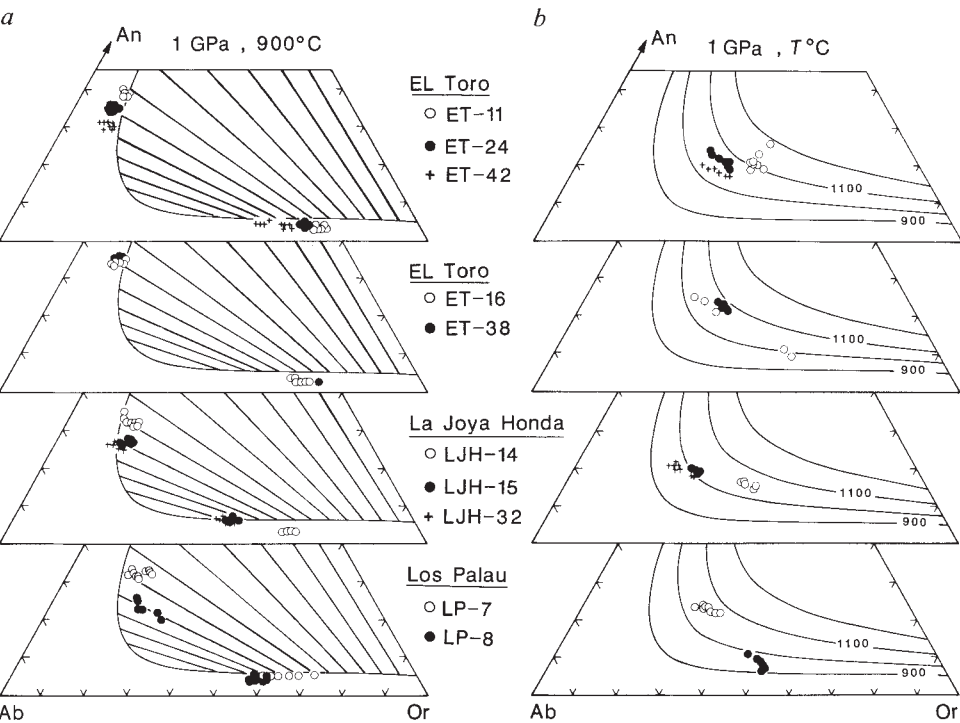


FIG. 1 *a*, Exsolved feldspar compositions, for each xenolith, plotted on stacked anorthite (An)-albite (Ab)-orthoclase (Or) diagrams. Symbols correspond to sample numbers as shown. Isotherm and tie lines for 900 °C are shown at 1.0 GPa (ref. 17). *b*, Reintegrated feldspar compositions plotted as in *a*. Isotherms are shown at 1.0 GPa for 900, 1,000, 1,100 and 1,200 °C (ref. 17).

equilibrated with homogeneous ternary feldspars. Diffusion through the garnet would be required to permit re-equilibration of exsolved plagioclase compositions with matrix sillimanite.

The paragneiss garnets are homogeneous, with almandine ranging from 49 to 55 mol %, pyrope from 39 to 47 mol % and grossular from 3 to 5 mol % (Table 2). Garnets contain abundant inclusions of quartz, sillimanite and rutile. If the garnet in each sample equilibrated with the reintegrated ternary feldspar, sillimanite ($X_{\text{Al}_2\text{SiO}_5} = 0.99$), rutile ($X_{\text{TiO}_2} = 0.97\text{--}0.98$), ilmenite ($X_{\text{FeTiO}_3} = 0.81\text{--}0.98$) and quartz, barometry may be applied using the feldspar activity/composition relations of ref. 17. For garnet we have used the mixing model of Anovitz and Essene²⁴. An ideal-solution model was used for rutile, sillimanite and ilmenite, and the activity of quartz was assumed to be unity. The garnet-sillimanite-quartz-plagioclase barometer⁸ yields pressures of 0.9–1.4 GPa at $T = 950\text{--}1,200^\circ\text{C}$; the garnet-rutile-ilmenite-plagioclase-quartz barometer⁹ also gives pressures of $\sim 0.9\text{--}1.3$ GPa. If lower pressure limits from the five samples that lack ilmenite are excluded, pressures of 0.9–1.3 GPa are obtained with the garnet-rutile-sillimanite-ilmenite-quartz barometer¹⁰. Pressures deduced from the three barometers correspond to depths of $\sim 33\text{--}48$ km (depending on temperature and the average specific gravity of the crustal rock column, chosen here to be 2.7). These depths are consistent with the estimated crustal thickness (35–45 km) in the study area^{25–27}, and show that the xenoliths sample the region of the crust/mantle boundary in central Mexico.

The mesoperthites provide evidence for regional metamorphism at $T > 950\text{--}1,125^\circ\text{C}$, higher than any temperatures recorded previously in regionally metamorphosed rocks of undoubted sedimentary parentage. The metamorphism spanned an area of at least 100 km^2 in central Mexico and may be related to the occurrence of similar metapelitic xenoliths 1,600 km to the north at Kilbourne Hole in New Mexico in the United States. We conclude that the lowermost crust of central Mexico was heated to $\sim 1,100^\circ\text{C}$ by underplated basaltic magmas^{5,28}. The base of the crust was subsequently cooled to $\sim 900^\circ\text{C}$, as recorded by the exsolved feldspars. We believe the xenoliths never equilibrated below 900°C and were excavated from nearly the depth they record by ascending basanitic magmas. Such extreme

quench temperatures ($800\text{--}950^\circ\text{C}$) for the feldspars suggest recent granulite facies metamorphism of the base of the crust in central Mexico; a mid-Tertiary underplating event would be consistent with moderate cooling rates of $5\text{--}10^\circ\text{C Myr}^{-1}$ for a regional terrane. These inferences are consistent with data in a chronological study of Kilbourne Hole metapelites, in which a metamorphic age of 1.6 Gyr was obtained from Rb–Sr isochrons of adjacent layers and an age of $34 (\pm 14)$ Myr from mineral isochrons²⁹.

We have attempted to unravel the cooling history of the xenoliths using data on the coarsening of exsolution lamellae in alkali feldspars³⁰, and the homogenization of plagioclases³¹. Data from coarsening experiments³⁰ indicate that lamellae of width $10\text{--}30\text{ }\mu\text{m}$ could be produced in 10^5 yr at 800°C , and 10^{-3} yr at $1,100^\circ\text{C}$ (with an initial lamella width of $0\text{ }\mu\text{m}$), but calculated homogenization rates³¹ are orders of magnitude slower. At $1,100^\circ\text{C}$, lamellae $10\text{ }\mu\text{m}$ wide would be homogenized in $10^{4\pm 1.5}$ yr, and $\sim 10^{10\pm 1}$ yr are needed for $10\text{-}\mu\text{m}$ -wide lamellae at 800°C . Although the homogenization rates, for which experiments involved coupled NaSi–CaAl diffusion³¹, are consistent with the preservation of feldspar lamellae in the xenoliths, the coarsening estimates are in sharp disagreement with our hypothesis of slow cooling subsequent to a mid-Tertiary regional metamorphic event. Perhaps the strongest evidence against a mesoperthite origin related to the host basalt is the absence of coarse perthites in volcanic rocks; to our knowledge all volcanic perthite occurrences are cryptoperthites/micropertthites with lamellae that are $\leq 1\text{ }\mu\text{m}$ wide. We believe that available experiments on lamellar coarsening are not relevant to ternary mesoperthites because the experiments were done on cryptoperthites with coherent structures at a pressure of 10^5 Pa (1 atm) and at temperatures $< 600^\circ\text{C}$, and, most importantly, they were done on binary alkali feldspars. Although higher temperatures should enhance coarsening, coupled Al/Si diffusion, required for exsolution of ternary feldspars, will greatly hinder coarsening^{30–32}. As the Mexican mesoperthites exsolved on the strain-free solvus with non-coherent structures, a mechanism involving nucleation and growth is required—a slower process than spinodal decomposition. Coarsening rates calculated from existing experimental data are inconsistent with



FIG. 2 Back-scattered electron image of a typical mesoperthite (sample ET24) showing the regular and continuous nature of exsolved plagioclase lamellae (darker bands). Quartz (dark) and rutile (light) are also present. Scale bar = $150\text{ }\mu\text{m}$.

geological evidence for the growth rates of mesoperthites with strongly ternary compositions^{33,34}.

Using depths of 40–45 km (corresponding to 1.1–1.2 GPa) with temperatures of 1,100–1,150 °C, for the peak of regional metamorphism, and present-day temperatures of 800–950 °C for the partially cooled, basal crustal rocks, average geothermal gradients were 26–28 °C km⁻¹ during the hypothesized basaltic underplating event and should be 20–21 °C km⁻¹ in central Mexico today. Our gradients compare favourably with those

estimated from measurements of heat flow obtained ~100 km to the north³⁵. We believe that our conclusions apply for the region extending from the central Mexican plateau to at least as far north as southern New Mexico (USA), where similar metapelite xenoliths occur at Kilbourne Hole^{36,37} and Engle Basin³⁸. Furthermore, as basaltic magmatism is still occurring in this region, deep-crustal metamorphism may be happening in central Mexico today. □

Received 26 June; accepted 6 October 1989.

- Griffin, W. L. & O'Reilly, S. Y. in *Mantle Xenoliths* (ed. Nixon, P.) 413–432 (Wiley, New York, 1987).
- Kay, R. W. & Kay, S. M. *Rev. Geophys. Space Phys.* **19**, 271–297 (1981).
- Roberts, S. J. & Ruiz, J. *J. geophys. Res.* **94**, 7961–7974 (1989).
- Ruiz, J., Roberts, S. J. & Patchett, P. J. *Terra Cogn.* **8**, 271 (1988).
- Bohlen, S. R. & Mezger, K. *Science* **244**, 326–329 (1989).
- Nixon, P. H. (ed.) *Mantle Xenoliths* (Wiley, New York, 1987).
- Essene, E. J. in *Evolution of Metamorphic Belts* (eds Daly, J. S., Cliff, R. A. & Yardley, B. W. D.) (Geol. Soc. Spec. Publ. No. 41, 1989).
- Kozio, A. M. & Newton, R. C. *Am. Miner.* **73**, 216–223 (1988).
- Bohlen, S. R. & Liotta, J. J. *J. Petrol.* **27**, 1025–1034 (1986).
- Bohlen, S. R., Wall, V. J. & Boettcher, A. L. *Am. Miner.* **68**, 1049–1058 (1983).
- Aranda-Gómez, J. J. thesis, Univ. Oregon (1982).
- Aranda-Gómez, J. J. & Ortega-Gutiérrez, F. in *Mantle Xenoliths* (ed. Nixon, P. H.) 75–84 (Wiley, New York, 1987).
- Luhr, J. F., Aranda-Gómez, J. J. & Pier, J. G. *J. geophys. Res.* **94**, 7916–7940 (1989).
- Stähle, V. *Earth planet. Sci. Lett.* **25**, 71–81 (1975).
- Padovani, E. L. R. & Carter, J. L. *Oregon Dept Geol. Miner. Ind. Bull.* **96**, 43–57 (1977).
- Bohlen, S. R. & Essene, E. J. *Contr. Miner. Petrol.* **62**, 153–169 (1977).
- Lindsley, D. H. & Nekvasil, H. *Eos* **70**, 506 (1989).
- Fuhrman, M. L. & Lindsley, D. H. *Am. Miner.* **73**, 201–215 (1988).
- Mora, C. I. & Valley, J. W. *Contr. Miner. Petrol.* **89**, 215–225 (1985).
- Powers, R. E. & Bohlen, S. R. *Contr. Miner. Petrol.* **90**, 401–409 (1985).
- Leyreloup, A., Bodinier, J. L., Dupuy, C. & Dostal, J. *Contr. Miner. Petrol.* **79**, 68–75 (1982).

- Robinson, J. V. thesis, Univ. Calif., Santa Cruz (1988).
- Padovani, E. L. R. thesis, Univ. Texas, Dallas (1977).
- Anovitz, L. M. & Essene, E. J. *J. Geol.* **95**, 633–645 (1987).
- Meyer, R. P., Steinhart, J. S. & Woollard, G. P. *Carnegie Inst. Wash. Publ. No. 622*, Ch. 5 (1961).
- Hales, A. L., Helsley, C. E. & Nation, J. B. *J. geophys. Res.* **75**, 7362–7381 (1970).
- Rivera, J. & Ponce, L. *Geofísica Int.* **25**, 383–402 (1986).
- Ruiz, J., Patchett, P. J. & Arculus, R. J. *Contr. Miner. Petrol.* **99**, 36–43 (1988).
- Padovani, E. R. & Reid, M. R. *New Mexico Bureau Mines Miner. Resources Mem.* **46**, 174–179 (1989).
- Yund, R. A. & Davidson, P. *Am. Miner.* **63**, 470–477 (1978).
- Grove, T. L., Baker, M. B. & Kinzler, R. J. *Geochim. cosmochim. Acta* **48**, 2113–2121 (1984).
- Mardon, D. & Yund, R. A. *Eos* **62**, 411 (1981).
- Brown, W. L., Becker, S. M. & Parsons, I. *Contr. Miner. Petrol.* **82**, 13–25 (1983).
- Parsons, I. & Brown, W. L. *Contr. Miner. Petrol.* **82**, 1–12 (1983).
- Smith, D. L. & Jones, R. L. in *Rio Grande Rift: Tectonics and Magmatism* (ed. Riecker, R. E.) 269–278 (Am. Geophys. Union, Washington, DC, 1979).
- Padovani, E. L. R. & Carter, J. L. *Geophys. Monogr. Ser.* **20**, 19–55 (1977).
- Bussod, G. Y. A. *Terra Cogn.* **2**, 266–267 (1982).
- Warren, R., Kudo, A. & Keil, K. in *Rio Grande Rift: Tectonics and Magmatism* (ed. Riecker, R. E.) 393–415 (Am. Geophys. Union, Washington, DC, 1979).

ACKNOWLEDGMENTS. We thank D. H. Lindsley and H. Nekvasil for providing calculations of their feldspar model in advance of publication, and Susan Fast for drafting Fig. 1. Critical reviews were provided by S. R. Bohlen, J. D. Clemens, M. A. Cosca, K. Mezger, R. C. Newton, R. Molina and H. N. Pollack. This study was supported by Sigma Xi, the University of Michigan Scott Turner Fund and the National Science Foundation.

Genetic relatedness in primitively eusocial wasps

Joan E. Strassmann*, Colin R. Hughes*,
David C. Queller*, Stefano Turillazzi†,
Rita Cervo†, Scott K. Davis§ & Keith F. Goodnight*

* Department of Ecology and Evolutionary Biology, Rice University,
Houston, Texas 77005, USA

† Dipartimento di Biologia Animale e Genetica, Università di Firenze,
Via Romana 17, 50125 Firenze, Italy

§ Department of Animal Science, Texas A&M University, College Station,
Texas 77843, USA

SOCIAL insects are characterized by indirect reproduction, in which most individuals achieve genetic success by helping to rear the offspring of colony mates. The evolution of such systems is expected to be sensitive to the genetic relatedness of colony mates. In general, reproductive altruism evolves most easily when relatedness is high, although it could be maintained under regimes of low relatedness after morphologically distinct castes have evolved^{1,2}. Most social insects belong to the order Hymenoptera, and are therefore haplodiploid (males haploid, females diploid); this genetic system may favour the evolution of altruism because it makes rearing a full sister ($r = \frac{3}{4}$) genetically more valuable than rearing one's own offspring ($r = \frac{1}{2}$) (refs 1–3). Here we report new estimates of relatedness from 14 species of polistine wasps lacking morphological castes. Female colony mates are often not full sisters and therefore seldom realize the full advantage made possible by haplodiploidy. But relatedness is always fairly high, in striking contrast to the situation in some species with morphological castes.

Methods for using genetic markers to estimate relatedness, with standard errors, have been available for a decade^{4–8}. But they have been applied to only three species of primitively eusocial insects (a halictid bee^{9,10}, an anthophorid bee¹¹ and a sphecoid wasp¹²), whose lack of morphologically distinct castes

makes them better than highly eusocial species as models for the early stages of social evolution. These methods have not been applied to primitively eusocial polistine wasps, which are the best-studied group. Colony mates in this group are known to be related from observational studies^{13–16} and from early electrophoretic methods that did not permit estimation of standard errors^{17,18}. The estimates reported here for 14 species of *Polistes* and *Mischocyttarus* confirm this finding, and in addition are sufficiently numerous and sufficiently well defined to allow some general tests of predictions about the role of relatedness in social insect societies.

Adult female wasps were collected from their nests at field sites in Texas, Yucatán (Mexico), Venezuela, and Italy. They were brought back to the laboratory and subjected to starch gel electrophoresis of proteins (see ref. 19 for methods). Between 1 and 6 variable proteins were found and used for estimation of relatedness⁸. The estimated relatedness values among female nestmates are generally high (Table 1), with a mean of 0.542 for the 15 populations (two geographically separated and genetically distinct populations of *Polistes exclamans* were included). This is nearly the same as relatedness to offspring, so on average, these species cannot gain much of a relatedness advantage from helping on their natal nest. The highest point estimates are very near the full-sister value of $\frac{3}{4}$, and the lowest attain only about half that value. The 95% confidence intervals for eight species include $\frac{3}{4}$, while eleven include $\frac{1}{2}$ (Fig. 1). Only three species (*P. gallicus*, *M. immarginatus*, *P. carolinus*) should be viewed as being significantly higher than $\frac{1}{2}$, and one (*P. annularis*) is significantly lower.

Clearly, relatednesses higher than the offspring value of $\frac{1}{2}$ cannot provide the whole explanation for helping. First, few species have an average r significantly above $\frac{1}{2}$ and one is significantly below. Second, in species with an average r near $\frac{1}{2}$, some colonies must be below this average. Third, including males would presumably lower the average as haplodiploidy does not increase sister-brother relatedness. Finally, the relatedness values that are relevant to some altruistic behaviours should be lower than those reported here. For example, autumn females

TABLE 1 Coefficients of relatedness (r) for *Mischocyttarus* and *Polistes* females

Species	Collection	Date	Sample sizes			$r \pm \text{s.e.}$
			colonies	females	loci	
<i>M. basimacula</i>	Yucatán	Dec. 1987	16	75	2	0.435 ± 0.116
<i>M. immarginatus</i>	Yucatán	Dec. 1987	19	123	3	0.765 ± 0.036
<i>P. versicolor</i>	Venezuela	July 1988	18	107	4	0.371 ± 0.084
<i>P. canadensis</i>	Venezuela	July 1988	20	129	3	0.339 ± 0.097
<i>P. annularis</i>	Texas	Oct. 1986	34	156	2	0.306 ± 0.075
<i>P. instabilis</i>	Texas	Oct. 1986	42	227	2	0.525 ± 0.064
<i>P. exclamans</i> 1	South Texas	Oct. 1986	46	312	1	0.557 ± 0.085
<i>P. exclamans</i> 2	East Texas	Oct. 1986	33	277	3	0.692 ± 0.104
<i>P. metricus</i>	Texas	Oct. 1986	19	145	5	0.567 ± 0.087
<i>P. carolinus</i>	Texas	Aug. 1984	16	59	6	0.630 ± 0.059
<i>P. dorsalis</i>	Texas	Oct. 1986	21	204	4	0.607 ± 0.079
<i>P. bellicosus</i>	Texas	Oct. 1986	31	289	1	0.338 ± 0.148
<i>P. dominulus</i>	Italy	Sept. 1988	8	57	4	0.652 ± 0.088
<i>P. gallicus</i>	Italy	Sept. 1988	12	134	2	0.803 ± 0.150
<i>P. nimpha</i>	Italy	Sept. 1988	10	103	3	0.540 ± 0.162

Sampling for each species was generally restricted to small (<6 ha) areas. When this was not true, data were combined from several small areas when an F_{ST} close to zero indicated homogeneity, or treated separately when there was significant differentiation (*P. exclamans*).

of some species will overwinter and serve as helpers to one of their natal nestmates who was not the queen of the natal nest^{13,14,20}, so their relatedness to the brood they help rear is only half of their relatedness to the nestmate who becomes the queen.

When relatedness to colony mates is less than $\frac{1}{2}$, there are two possible explanations for the selective maintenance of eusociality. The first is that individuals may be able to direct their aid towards colony mates who are especially closely related, perhaps rearing full sisters even when most colony mates are half-sisters. Such discrimination has been demonstrated in honeybees^{21,22}, though the effect is generally weak. It requires that individuals use cues learned from their own phenotypes. This seems unlikely for *Polistes* because adults isolated from their nest subsequently fail to distinguish kin from non-kin²³. These experiments suggest that kin-recognition cues are derived from the nest or brood, sources that are unlikely to carry information useful for within-colony discriminations. Thus, the available evidence seems to support the second possible explanation, which is that reproductive altruism may evolve even when $r < \frac{1}{2}$. According to inclusive fitness theory^{1,2}, helping is favoured when $rb > \frac{1}{2}c$, where b is the fitness benefit to the individual helped, c is the cost to the helper, r is relatedness to the brood reared, and $\frac{1}{2}$ is the relatedness of the helper to its own offspring. So

even when $r < \frac{1}{2}$, altruism can be favoured when the benefit is sufficiently higher than the cost. Several studies have shown that benefits may exceed costs in *Polistes*²⁴⁻²⁷.

Although the data do not support the prediction that individuals usually gain a relatedness advantage by rearing offspring that are not their own, they do confirm some of Hamilton's ideas^{1,2}. With the addition of our 14 species to the 3 previously studied primitively eusocial species⁹⁻¹², there are sufficient data to support the generalization that relatedness is always fairly high in species without morphological castes. This permits us to draw an interesting contrast to species with castes. In five species of social wasps with at least some degree of caste development, relatedness takes on lower values, ranging from 0.11 to 0.40 (refs 19, 28). For ants with highly specialized worker castes, relatedness within colonies is variable, but is extremely low in some species²⁹⁻³¹. Thus when individuals are constrained by a specialized worker morphology they may accept a subordinate, nonreproductive role to aid individuals that may not be highly related. However, as expected from inclusive fitness theory, to the extent that individuals are unconstrained in their choices, they perform such sacrifices only for rather close relatives. □

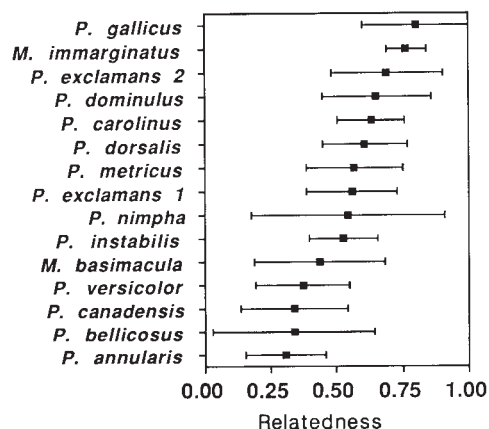


FIG. 1 Point estimates and 95% confidence intervals for relatedness among female nestmates in *Polistes* and *Mischocyttarus*. Confidence intervals are calculated using values of the t -distribution with degrees of freedom equal to one less than the number of colonies.

Received 26 July; accepted 10 October 1989.

- Hamilton, W. D. *J. theor. Biol.* **7**, 1-52 (1964).
- Hamilton, W. D. *A. Rev. Ecol. Syst.* **3**, 193-232 (1972).
- Trivers, R. L. & Hare, H. *Science* **191**, 249-263 (1976).
- Craig, R. & Crozier, R. H. *Evolution* **33**, 335-341 (1979).
- Pamilo, P. & Varvio-Aho, S.-L. *Behav. Ecol. Sociobiol.* **6**, 91-98 (1979).
- Pamilo, P. & Crozier, R. H. *Theor. Popul. Biol.* **21**, 171-193 (1982).
- Pamilo, P. *Genetics* **107**, 307-320 (1984).
- Queller, D. C. & Goodnight, K. F. *Evolution* **43**, 258-275 (1989).
- Crozier, R. H., Smith, B. H. & Crozier, Y. C. *Evolution* **41**, 902-910 (1987).
- Kukuk, P. F. in *Genetics of Social Evolution* (eds Breed, M. D. & Page, R. E.) 183-202 (Westview, Boulder, 1989).
- Schwarz, M. P. *Behav. Ecol. Sociobiol.* **21**, 387-392 (1987).
- Ross, K. G. & Matthews, R. W. *Amer. Nat.* (in the press).
- Pardi, L. *Boll. Ist. Entom. Univ. Bologna* **15**, 25-84 (1942).
- West-Eberhard, M. J. *Misc. Publ. Mus. Zool. Univ. Mich.* **140**, 1-101 (1969).
- Little, M. *Behav. Ecol. Sociobiol.* **2**, 229-246 (1977).
- Strassmann, J. E. *Z. Tierpsychol.* **69**, 141-148 (1985).
- Strassmann, R. A. & Whitt, G. S. *Behav. Ecol. Sociobiol.* **2**, 339-351 (1977).
- Metcalfe, R. A. & Whitt, G. S. *Behav. Ecol. Sociobiol.* **117**, 147-176 (1981).
- Lester, L. J. & Selander, R. K. *Amer. Nat.* **117**, 147-176 (1981).
- Queller, D. C., Strassmann, J. E. & Hughes, C. R. *Science* **242**, 1155-1157 (1988).
- Strassmann, J. E. *Science* **204**, 207-209 (1979).
- Frumhoff, P. C. & Schneider, S. *Anim. Behav.* **35**, 255-262 (1986).
- Page, R. E., Robinson, G. E. & Fondrk, M. K. *Nature* **338**, 576-579 (1989).
- Gamboa, G. J., Reeve, H. K. & Pfennig, D. W. *A. Rev. Entomol.* **31**, 431-454 (1986).
- Metcalfe, R. A. & Whitt, G. S. *Behav. Ecol. Sociobiol.* **2**, 353-360 (1977).
- Noonan, K. M. in *Natural Selection and Social Behavior* (eds Alexander, R. D. & Tinkle, D. W.) 18-44 (Chiron, New York, 1981).
- Strassmann, J. E., Queller, D. C. & Hughes, C. R. *Ecology* **69**, 1497-1505 (1988).
- Queller, D. C. *Proc. natn. Acad. Sci. U.S.A.* **86**, 3224-3226 (1989).
- Ross, K. G. *Nature* **323**, 398-400 (1986).
- Pamilo, P. *Heredity* **48**, 95-106 (1982).

30. Pearson, B. *Behav. Ecol. Sociobiol.* **12**, 1–4 (1983).
 31. Ross, K. G. & Fletcher, D. J. C. *Behav. Ecol. Sociobiol.* **17**, 349–356 (1985).

ACKNOWLEDGEMENTS. We thank J. M. Carpenter for identifying some of the wasps, J. González and A. Fernández for assistance in Venezuela, A. R. Templeton and T. Bertin for access to laboratory facilities, and A. Treves, J. Negrón and C. R. Solís for help with data collection. For access to field sites we thank Houston Shell and Concrete, Brazos Bend State Park and Universidad Central de Venezuela in Maracay. Supported by the NSF, National Geographical Society and the John Simon Guggenheim Foundation.

Telemetered *in vivo* strain analysis of locomotor mechanics of brachiating gibbons

Sharon M. Swartz*†‡, John E. A. Bertram§
& Andrew A. Biewener*§

* Committee on Evolutionary Biology and §Department of Organismal Biology and Anatomy, The University of Chicago, 1025 East 57th Street, Chicago, Illinois 60637, USA

† Yerkes Regional Primate Research Center, Emory University, Atlanta, Georgia 30322, USA

THE slender elongated form that is characteristic of the forelimb long bones of gibbons (*Hylobates*) has long been attributed to their functional adaptation to habitual armswinging locomotion^{1–3}, although potential selective advantages of this morphology for brachiation have yet to be demonstrated. If the forces exerted on the limb skeleton during brachiation indeed differ greatly from those of other locomotor modes, then the changes in skeletal loading accompanying a shift in locomotor behaviour could favour alterations in skeletal morphology in brachiating lineages. *In vivo* skeletal strain patterns recorded by using radiotelemetry during brachiation indicate that the forelimb bones of the gibbon are loaded in substantial tension and show reduced bending and compression in comparison with those of other mammals. We suggest that this unique loading regime could have contributed to the evolution of the distinctive morphology of hylobatid limbs.

Theoretical analyses of the mechanics of brachiation have modelled gibbon locomotion after ideal pendula^{4,5}; these models predict that the supporting limb is subjected to axial tension under the load exerted by gravity acting on the animal's mass. Correspondingly, it has been suggested that tensile loading is the key feature distinguishing brachiation from walking, running and jumping^{6,7}. However, models of the response of bone to mechanical loading have argued against the possibility that bones are ever regularly subjected to net tension⁸. Furthermore, because all vertebrate long bones so far studied experience primarily bending and compressive forces^{9–12}, axial tensile loading is a novel loading regime for a limb. This would alter the criteria for natural selection acting on skeletal design, given that a structure's ability to resist tensile, but not compressive and bending forces, is independent of both length and cross-sectional shape, and that bending, particularly due to eccentric loads, increases in proportion to length¹³. If, in brachiation, bone compression and bending are minimal and tension predominates, an increased bone length will not be structurally disadvantageous.

Tensile loading may not characterize suspensory support, however, because skeletal loading depends not only on external forces but also on internally transmitted, largely compressive muscle forces. To resolve this issue, we recorded *in vivo* surface bone strains at the dorsal and ventral midshaft of the humerus, radius and ulna of four freely brachiating white-handed gibbons (*Hylobates lar*) by using radiotelemetry¹⁴. Principal strain mag-

nitudes and their orientations with respect to the bone's long axis were determined for each rosette strain gauge, and used to characterize bone loading¹⁵.

For all forelimb long bones, principal strain magnitude peaked at midsupport phase of the swing, as the animal's body reached the bottom of the arc swept out by the forelimb; near-zero strains were obtained when the limb was not grasping a support (Fig. 1). Considerable variability in strain pattern and peak magnitude among swings was observed, contrasting with the stereotypic patterns recorded in walking and running animals^{9,11,16}. This seems to be due both to the unrestricted movement in a large enclosure compared with controlled treadmill locomotion, and to development of localized strain from muscles attaching near the gauges.

The largest principal strains recorded on the dorsal and ventral surfaces of the ulna were tensile, similar in magnitude, and oriented close to the bone's longitudinal axis, indicating overall axial tension (95% of total strain) with a small amount of bending (Fig. 2a; Table 1). In the radius, principal strains were again oriented longitudinally, and showed substantial compression on the ventral midshaft surface with smaller tensile strains dorsally (Fig. 2b; Table 1). Unlike the ulna, loading of the radius was characterized by dorso-ventral bending (83% of total strain) superimposed on axial compression. In the humerus, peak principal strains for both the dorsal and ventral surfaces were tensile but differed in magnitude, and strain orientation deviated 45–60° from the bone's axis (Fig. 2c; Table 1). The humerus thus experienced a combination of axial tension, torsion (indicated by the off-axis orientation) and bending.

The overall axial tension recorded from the ulna was similar (within 11%) to values expected for a bone functioning as a passive suspending cord of a pendulum¹⁴. Given this tensile loading and the predicted net tensile loading of the forelimb as a whole, the bending and compression of the radius may seem surprising. This loading disparity arises from differences between the two bones in morphology, transmitted muscular force, and weight-support roles. In long bones, the main source of bending may be axial force exerted about the longitudinal curvature of the bone¹⁷; in the gibbon, radii are seven times more curved than ulnae¹⁴. Muscle force is typically the greatest constituent of axial force exerted on long bones (generally two to three times the magnitude of the reaction force of the body acting against the ground⁹). Gibbon forearm flexor muscles are positioned ventrally so that their pull will bend the bone in a ventrally concave fashion, dictated by the radius's dorso-ventral curvature, and during the support phase of brachiation, flexor activity is marked^{18–20}. Additionally, weight is transferred in tension directly from the humerus to the ulna through the interlocking humero-ulnar joint, whereas the direct connections from humerus to radius are exclusively ligamentous, and the interosseus membrane joining the two forearm bones is not stiff

TABLE 1 Mean peak maximum principal strain magnitudes and orientations

	$\mu\epsilon$	(s.d.)	Orientation	(s.d.)
Ventral ulna	+1,421	(337)	2°	5°
Dorsal ulna	+1,270	(167)	6°	1°
Ventral radius	–1,638	(453)	1°	2°
Dorsal radius	+746	(248)	5°	4°
Ventral humerus	+419	(180)	–44°	3°
Dorsal humerus	+1,492	(71)	+62°	4°

Mean peak-maximum principal strain magnitudes (in microstrain, $\mu\epsilon$) and orientations (in degrees deviation from the longitudinal axis of the bone) for each recording site. Tensile strains are positive values, compressive strains negative. Orientations are positive if they deviate from the bone's long axis in the proximo-medial direction, negative if they deviate in the proximo-lateral direction. The s.d. of the values for each site are given in parentheses after the mean value.

† Present address: Departments of Cell Biology and Anatomy and Anthropology, Northwestern University, 303 East Chicago Avenue, Chicago, Illinois 60611, USA.

or thick enough to transfer substantial stress between them. Compression of the radius indicates both that muscle force is preferentially exerted through the radius rather than through the ulna, and that this force provides the primary loading of the radius, whereas the animal's body weight is transferred from the humerus to the hand through the ulna.

The lesser tensile strain recorded from the humerus compared with that recorded from the ulna probably results from com-

pressive forces exerted by the arm and shoulder musculature¹⁸⁻²⁰ counteracting tension from suspensory weight support, and the six-times greater cross-sectional area of the humerus¹⁴. The curvature of the humerus and the differential muscle activity on the flexor and extensor aspects of the arm probably produces the bending we observed. The significant torsion of the humerus reflects the bone's resistance to the lateral rotation of the gibbon's body about the long axis of the limb during the swing²¹. During this movement, the humerus rotates little at the shoulder joint and the supporting hand remains stationary²². Thus, free rotation of the radius and proximal carpus about the humerus and distal carpus provides propulsive motion^{22,23} while freeing the forearm bones from torsion. Stabilization at the shoulder combined with the hinge-like morphology of the humero-ulnar joint can then result in torsional loading of the humerus.

Our results show that some bones can be regularly loaded in net tension. Hence, models of adaptive bone remodelling positing bone resorption under tensile loading must be revised. Our results also indicate a new explanation for the elongation of the forelimb skeleton in gibbons that is based on the reduction of compressive and bending loads inherent in suspensory support. In general, limb elongation is advantageous in terms of increased stride length, velocity and greater reach, especially in grasping primates. However, greater limb-bone length is probably constrained in most terrestrial mammals partly by the risks engendered by excessive bending and the associated large strains when long slender bones experience off-axis loading. In contrast to terrestrial locomotion, in which external loading (weight support) and muscle forces are additive in generating compressive

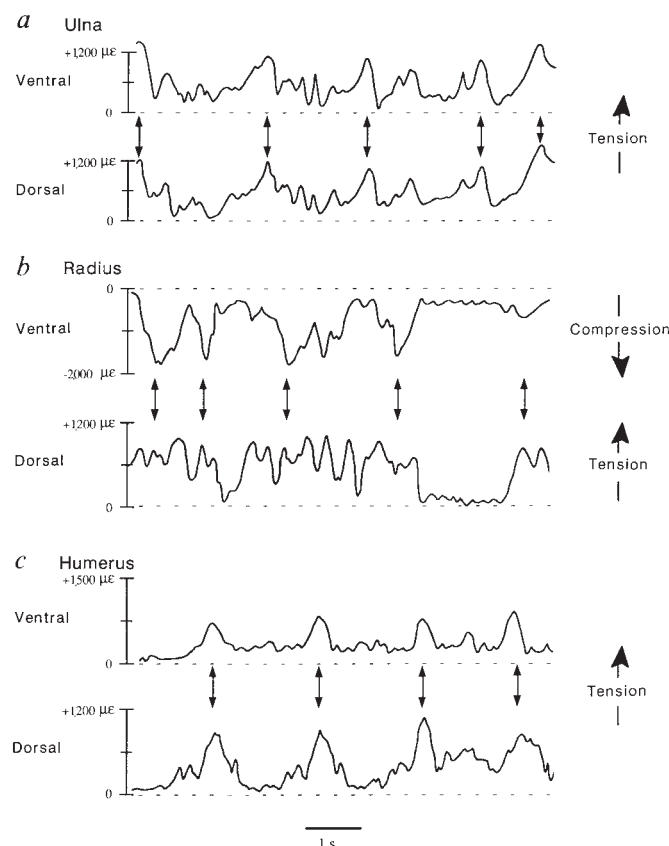


FIG. 1 Representative principal strain recordings from the ventral and dorsal midshaft surfaces of the ulna (a), radius (b) and humerus (c) of one gibbon for a duration of 8 s, correlated with swing kinematics based on light films (60 frames per second) taken during recordings. The mid-support phase of each swing is indicated by an upward pointing arrow. Strain magnitudes are in microstrain (strain $\times 10^{-6}$). Tensile strains are positive and compressive strains are negative. Bone strains reached their greatest absolute values at midswing, and strain values return close to zero as the limb is swung forward (protracted) during the nonsupporting phase of the locomotor cycle (analogous to the swing phase of terrestrial locomotion). All recordings were made at the Yerkes Regional Primate Research Center, Emory University, Atlanta, Georgia. By using aseptic surgery under general anaesthesia, we attached rosette strain gauges with a self-catalysing cyanoacrylate adhesive to the dorsal and ventral midshaft cortices of the right humerus and left radius and ulna at the bones' diaphyses, as near as possible to midshaft level but in areas free of direct muscle attachment to minimize localized distortions in overall bone strain due to muscle pull. The procedure was similar to that previously described²⁴. Lead wires were led subcutaneously up the arm and over the shoulder, externalized through the skin between the scapulae at approximately the level of the third thoracic vertebra, and soldered to a connector; all lead wires were left slack to permit rapid and forceful limb movements without dislodging the gauges. Following 2-3 days post-operative recovery, *in vivo* strain data were then recorded over a 2-3 day period using radiotelemetry. Each gibbon carried a strain transmitter on its back as it brachiated freely in a large ($9 \times 2 \times 3$ m) outdoor enclosure. The eight bridge channels were multiplexed to a single c. annel telemetry transmitter (Transkinetics TXM-205; frequency response, 35 kHz; transmission range, 50 m) which was housed in the pocket of a vest-style harness that fitted snugly around each animal's chest and back. Each channel was sampled at a frequency of 125 Hz. After 2-3 days of recordings, all gauges and wires were removed from the animals.

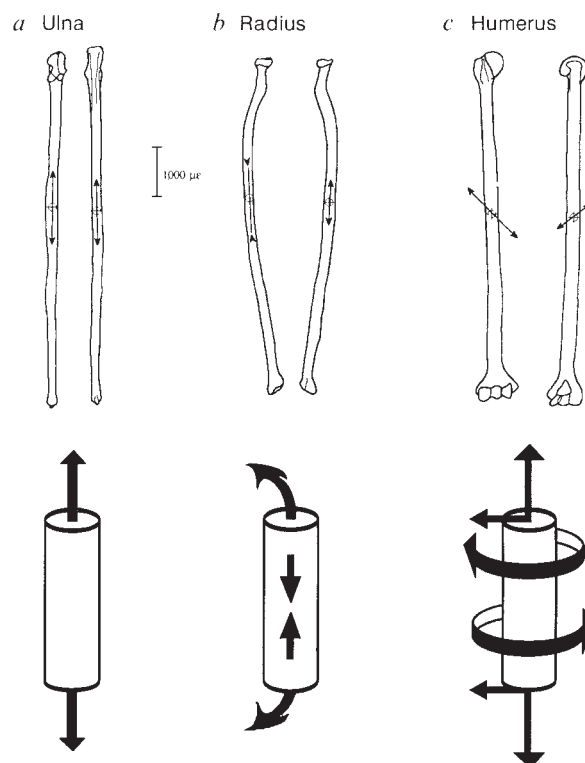


FIG. 2 Diagrammatic views of the ventral (left) and dorsal (right) ulna (A), radius (B) and humerus (C) with strain-gauge attachment sites indicated by the cross-hatched circles. Vectors depicting mean peak principal-strain magnitude and orientation with respect to the bone's longitudinal axis as recorded from the implanted strain gauges are indicated at each gauge attachment site; outwardly directed arrows indicate tension, whereas inwardly directed arrows indicate compression. The cylindrical model below each bone schematically shows each bone's overall loading mode: ulna, axial tension; radius, compression and dorsoventral bending; and humerus, tension, dorsoventral bending and torsion.

and bending stresses, muscle forces (compressive) and external weight support (tensile) will oppose one another during brachiation, reducing net axial strain in the skeleton. Brachiation further reduces bending of the forelimb skeleton by suspending the animal's weight directly in line with the bones' long axes at the time in the swing cycle when external forces are greatest¹⁴. By reducing bending, selection for increased bone length in brachiators could be favoured with minimal risk or increased strain and hence bone failure. □

Received 1 July; accepted 5 October 1989.

- Keith, A. *J. Anat.* **37**, 18–40 (1903).
- Andrews, P. & Groves, C. E. in *Gibbon and Siamang* (ed. Rumbaugh, D. M.) 167–218 (Karger, Basel, 1976).
- Preuschoft, H. & Demes, B. in *Size and Scaling in Primate Biology* (ed. Jungers, W. L.) 383–399 (Plenum, New York, 1985).
- Fleagle, J. G. *Nature* **248**, 259–260 (1974).
- Preuschoft, H. & Demes, B. in *The Lesser Apes: Evolutionary and Behavioral Biology* (eds Preuschoft, H., Chivers, D. J., Brockelman, W. V. & Creel, N.) (Edinburgh University Press, 1984).
- Kummer, B. *Anthrop. Anz.* **32**, 74–82 (1970).
- Oxnard, C. E. *The Order of Man* (Yale University Press, New Haven, 1984).
- Currey, J. D. *The Mechanical Adaptations of Bones* (Princeton University Press, New Jersey, 1984).
- Biewener, A. A., Thomason, J., Goodship, A. E. & Lanyon, L. E. *J. Biomech.* **16**, 565–576 (1983).
- Lanyon, L. E. & Bourne, S. *J. Bone Jt Surg.* **A61**, 263–273 (1979).
- Biewener, A. A., Swartz, S. M. & Bertram, J. E. A. *Calcif. Tissue Int.* **39**, 390–395 (1986).
- Biewener, A. A. & Taylor, C. R. *J. exp. Biol.* **123**, 383–400 (1986).
- Wainwright, S. A., Biggs, W. D., Currey, J. D. & Gosline, J. M. *Mechanical Design in Organisms* (Wiley, New York, 1976).
- Swartz, S. M. *The Biomechanics and Structural Design of the Forelimb of Brachiating Primates* (The University of Chicago Press, 1988).
- Dally, J. W. & Riley, W. F. *Experimental Stress Analysis* (McGraw-Hill, 1978).
- Rubin, C. T. & Lanyon, L. E. *J. exp. Biol.* **101**, 187–212 (1982).
- Bertram, J. E. A. & Biewener, A. A. *J. theor. Biol.* **131**, 75–92 (1988).
- Jungers, W. L. & Stern, J. T. *Jr Science* **208**, 617–619 (1980).
- Jungers, W. L. & Stern, J. T. *Jr Int. J. Primat.* **2**, 18–33 (1981).
- Jungers, W. L. & Stern, J. T. Jr in *The Lesser Apes: Evolutionary and Behavioral Biology* (eds Preuschoft, H., Chivers, D. J., Brockelman, W. V. & Creel, N.) 119–134 (Edinburgh University Press, 1984).
- Avis, V. *Southw. J. Anthropol.* **18**, 119–148 (1962).
- Jenkins, F. A. Jr, Bombroski, P. J. & Gordon, E. P. *Am. J. phys. Anthropol.* **48**, 65–76 (1978).
- Jenkins, F. A. Jr *Symp. zool. Soc. Lond.* **48**, 429–451 (1981).
- Lanyon, L. E. & Baggot, D. G. *J. Bone Jt Surg.* **B58**, 436–443 (1976).

ACKNOWLEDGEMENTS. We thank Drs R. Tuttle and M. LaBarbera for comments on this manuscript. Funds were provided by the NSF Doctoral Dissertation Improvement Program, the Hinds Fund, University of Chicago; the Louis S. B. Leakey Foundation; the Eli A. Niernman Foundation; and the Searle Graduate Fellowship Program. We are particularly grateful to Mr Joseph Ting for designing and making the telemetry/multiplexor circuitry. We are also grateful for the cooperation of the Yerkes Regional Primate Research Center, in particular Dr Harold McClure, Dr Larry Byrd, Dr Jeremy Dahl, and Dr Brent Swenson and the Yerkes Center veterinary staff. Research at the Yerkes Center is supported by the NIH Animal Resources.

Changes in spinal cord reflexes after cross-anastomosis of cutaneous and muscle nerves in the adult rat

Stephen B McMahon & Patrick D Wall*

Department of Physiology, St Thomas' Hospital Medical School, Lambeth Palace Rd, London SE1 7EH, UK

* Cerebral Functions Research Group, University College London, Gower Street, London WC1E 6BT, UK

EVIDENCE exists that the specification of afferent nerves and their central connections in the embryo may depend in part on influences from the peripheral target innervated^{1–7}. We have now investigated whether such peripheral determination persists in the adult rat using the unmyelinated afferent system of C fibres, which differ chemically in the adult depending on their target^{8–11}. We have previously shown that if the cutaneous sural nerve and the muscle gastrocnemius nerve are cross-anastomosed so that they grow to each other's target, the A fibres establish functional endings and the C fibres change their chemistry to that which is appropriate for the new target^{9,12–14}. Here we report that in normal adult rats, a short train of stimuli to the cutaneous sural nerve produced a brief facilitation of the flexion reflex, lasting on average only 5 min, whereas similar stimulation of the gastrocnemius-muscle nerve enhanced this reflex for an average of 54 min. In cross-anastomosed animals, stimulation of the gastrocnemius nerve (innervating skin) induced a brief potentiation of the flexion reflex, lasting on average only 3 min. By contrast, stimulation of sural nerve (innervating muscle) produced a potentiation of this reflex lasting 57 min. Thus the ability of adult afferent nerves to potentiate the flexion reflex depends on the target with which they make contact. We propose that tissue-specific factors influence some of the central actions of primary afferent neurons in the adult.

In adult decerebrate low-spinal rats, with normal nerves or cross-anastomosed or self-anastomosed sural nerves (s.n.) and gastrocnemius nerves (made 10–14 weeks previously), the flexion reflex was measured as the discharge in small branches of the nerve to biceps femoris/semitendinosus containing a few motoneurons¹⁵. The flexion reflex was elicited by applying a standard 3-s suprathreshold pinch stimulus to the middle three

toes of the ipsilateral hindpaw, and the multiunit discharge was recorded and counted for the 3 s of the stimulus. The reflex was initially similar in all animals and consisted of a phasic high-frequency discharge lasting 0.5–2 s, and its magnitude was stable from trial to trial. Repeated determinations of the flexion reflex were made at 5-min intervals. After obtaining consistent control responses over at least a 20 min period, conditioning stimuli of 20 shocks at C-fibre strength (5 mA, 500 μ s, 1 Hz) were given to the normal or anastomosed nerves, and the subsequent changes in the size of the flexion reflex were determined at 1-min intervals for 5 min, and at 5-min intervals thereafter.

Conditioning the normal sural nerve increased the response to the standard pinch stimulus both in terms of the frequency of firing of motoneurons and duration for which they fired. The facilitation persisted for on average 5 ± 1.6 min ($\pm$ s.e.m.). The magnitude of the enhancement of the flexion reflex was typically 200–300%. Figure 1 shows the time course of change in one typical case.

When the same conditioning stimulus was applied to normal g.n. there was an enhancement of the flexion reflex lasting for an average of 54 ± 8.3 min. A typical example is shown in Fig. 1. The difference in duration of facilitations was highly significant ($P < 0.01$), and is similar to that previously reported for normal animals^{15,16}. The modulation induced by stimuli conditioning g.n. often appeared to have two phases: an early phase resem-

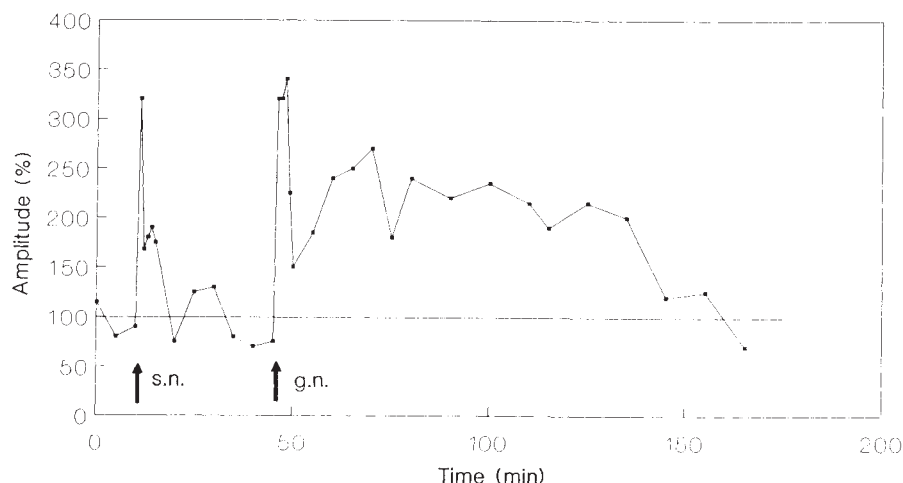
TABLE 1 Enhancement of flexion reflex induced by conditioning stimulus

Nerve stimulated	Average enhancement (min)	s.e.m.	No.
Normal s.n.	5	1.6	28
S.n. reinnervating skin	2	0.4	4
S.n. innervating muscle	59	9.4	5
Normal g.n.	54	8.3	16
G.n. reinnervating muscle	77	11.3	3
G.n. innervating skin	3	0.7	6

The duration of enhancement of flexion reflex (in min) induced by a conditioning stimulus (20 shocks, 5 mA, 500 μ s, 1 Hz) delivered either to the normal s.n. or g.n., or to s.n. and g.n. which had been cut and had reinnervated their own normal targets, or to s.n. and g.n. which had innervated inappropriate targets after cross-anastomosis.

FIG. 1 The time course of enhancement of the flexion reflex in one normal animal after a conditioning stimulus train to the s.n. (at a time indicated by arrow marked s.n.) and the g.n. (at a time indicated by arrow marked g.n.). The magnitude of the flexion reflex (the number of discharges evoked by the standard pinch stimulus) is plotted as a percentage of the average level determined in the control period. Note that whereas the s.n. produced only a short-lived facilitation, the g.n. produced a more prolonged facilitation.

METHODS. A small branch of the nerve to biceps femoris was dissected, typically containing 5–10 motoneurons activated by the pinch test stimuli. Multiunit efferent action potentials were recorded using conventional techniques with amplifier filters set at 1 kHz to 8 kHz, and counted with a window discriminator set to detect all voltage fluctuations greater than three times the noise level. The total number of counts occurring in the 3 s after the pinch stimuli are plotted here as a percentage of the counts obtained in the control period. The duration of facilitation was measured as the time



taken for the amplitude of the reflex to fall below 120% of the control level.

bling that seen with the conditioning of s.n. and a later phase which reached a maximum with a latency of 10–20 min. The order of presentation of conditioning stimuli did not affect the nature of facilitations.

Recordings on cross- or self-anastomosed nerves showed that the nerves had successfully innervated skin or muscle^{12,13}. The distal nerve trunk appeared to be of a normal size. We have previously shown that >70% of peripheral fibres successfully negotiate the anastomosis and regenerate to the inappropriate target^{12,13}.

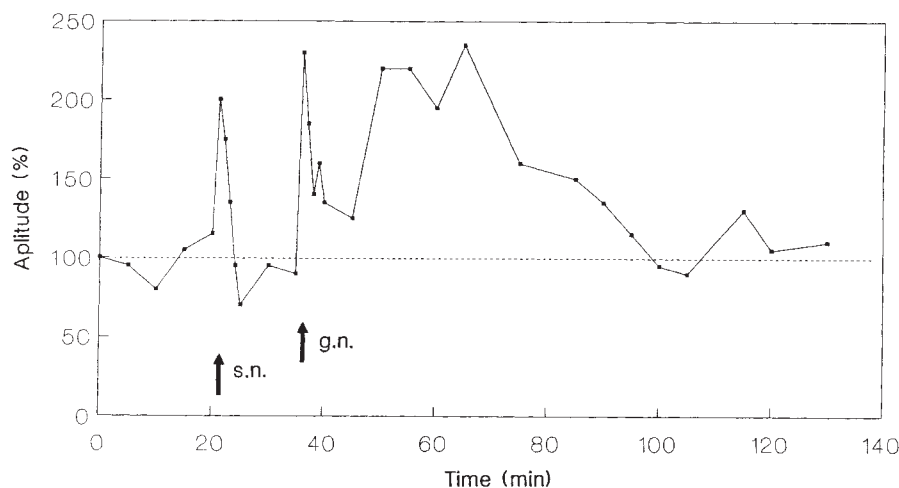
The conditioning stimuli produced very different changes in animals with the cross-anastomosed nerves. Stimuli to the g.n., then innervating skin, resulted in an enhancement of the flexion reflex, but this lasted only 3 ± 0.7 min (compared with 54 ± 8.3 min in normal animals, $P < 0.01$; Fig. 2). Stimulation of s.n., then innervating muscle, produced a long-lasting facilitation, starting within 1 or 2 min after application of the conditioning stimulus, and lasting 59 ± 9.4 min (compared with 5 min in normal animals, $P < 0.01$). These facilitations also tended to show an early and a late phase, like the effect seen in normal g.n. Table 1 shows the duration of the effects elicited by trains of conditioning stimuli to the different nerves.

For self-anastomosed nerves, the nature of the flexion reflex

and the facilitatory effects of conditioning the s.n. and g.n. were quantitatively similar to those in intact animals. That is, conditioning the s.n. enhanced the flexion reflex for 2 ± 0.4 min, and conditioning the g.n. enhanced the reflex for 77 ± 11.3 min ($P > 0.05$, compared with intact controls in both cases; Table 1). One typical example is shown in Fig. 3. The onset of facilitation was also similar, in all cases beginning 1–3 min after the conditioning stimulus. The only apparent difference between responses in intact and self-anastomosed nerves was in the magnitude of facilitation, which tended to be less in the latter group, typically 150–250%. Hence, the changes seen with cross-anastomosed nerves cannot be attributed to a nonspecific consequence of nerve section or regrowth.

To avoid the possibility of spread of electric current, conditioning by the injury discharge resulting from nerve section was used. Wall and Woolf¹⁶ noted that enhancement of the flexion reflex could be triggered by the injury discharge produced by section of peripheral nerves as well as by electrical stimulation. In normal animals, the section of s.n. induced a relatively short-lived enhancement of the flexion reflex, whereas that from the section of g.n. was much more prolonged ($P < 0.01$, see Table 2). Similar facilitations were seen with self-anastomosed nerves (Table 2). By contrast, in the animals with

FIG. 2 The time course of enhancement of the flexion reflex in one animal with cross-anastomoses of the s.n. and g.n. The figure is constructed in the same way as Fig. 1. Note that in this animal, in contrast to normal animals, the s.n. (innervating muscle) produced a long-lasting facilitation, whereas the g.n. does not.



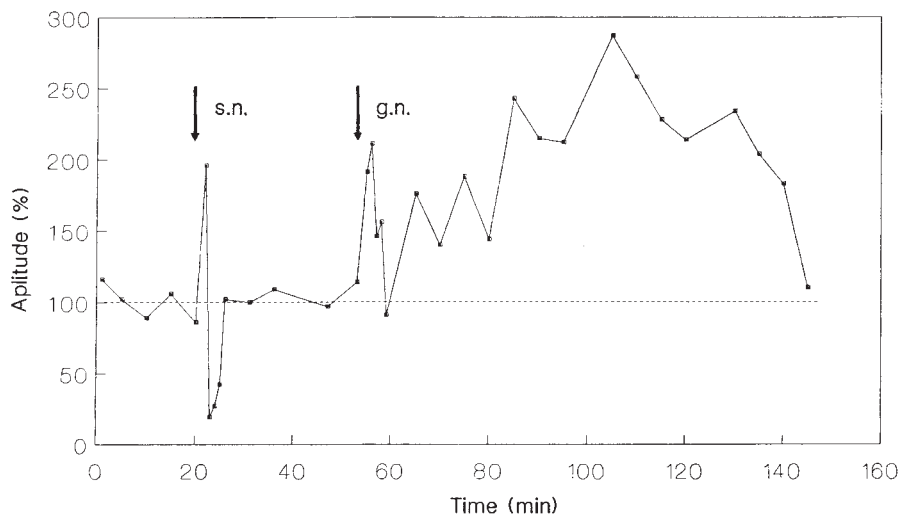


FIG. 3 The time course of enhancement of the flexion reflex in one animal with self-anastomoses of the s.n. and g.n. The figure is constructed in the same way as Fig. 1. Note that in this animal, like animals with intact nerves,

the g.n. (innervating muscle) produced a long-lasting facilitation, whereas the s.n. (innervating skin), does not.

cross-anastomoses, the section of s.n. (supplying muscle) elicited long-lasting effects, whereas the section of the g.n. (innervating skin), only short-lasting changes ($P<0.01$, Table 2).

The facilitation of the flexion reflex is known to depend on the activation of afferent C-fibers and to be due to an altered heterosynaptic excitability within the spinal cord^{16,17}. When the sural or gastrocnemius nerves reinnervate inappropriate tissue they change their normal characteristics and take on those appropriate for the tissue that they then innervate. That is, s.n. afferents innervating muscle induce long-lasting facilitation, whereas the g.n. afferents innervating skin lose this capacity. These changes in the spinal cord are presumably secondary to alterations in the properties of primary afferents when they innervate inappropriate tissue. The inappropriate tissue is presumably the skin or muscle itself, although it could be the non-neural elements in the nerves leading to the tissue. We have previously shown that such inappropriate reinnervation can alter the histochemistry of primary afferents. Normal s.n. C-fibres contain high levels of substance P and the enzyme FRAP (fluoride resistant acid phosphatase), whereas normal g.n. afferents do not; when these nerves are cross-anastomosed, the expression of these markers become appropriate for the tissue innervated^{9,14}. It could be that the altered ability to modulate the flexion reflex reflects some parallel change in the histochemistry of the primary afferents.

Thus, peripheral targets can not only regulate some properties

of primary afferents, but can extend this influence to a control of their central effects. In adult animals, isolation of afferents from their peripheral targets by axotomy can produce several plastic changes within the central nervous system, including a reorganization of the receptive fields of dorsal horn cells¹⁸⁻²¹. The strength of monosynaptic connections from Ia fibres may be reduced when these fibres inappropriately reinnervate non-spindle targets²². We have recently found²⁴ that when gastrocnemius afferents regenerate to skin, there is a dramatic change in the central responses after their stimulation. In normal animals, stimulation of unmyelinated g.n. afferents excites relatively few dorsal-horn cells. In animals with cross-anastomosed nerves, the central effect of gastrocnemius afferent barrages is greatly enhanced and many more dorsal-horn neurons are excited by these afferents. This too is not a non-specific effect due to axotomy or regrowth, but is dependent on some influence arising from the new tissue innervated. The changes we have reported here could similarly indicate the presence of tissue-specific trophic influences, the effects of which on primary afferents could further influence the establishment or maintenance of specific patterns of central responses. □

TABLE 2 Enhancement of flexion reflex induced by injury discharge

Nerve stimulated	Average enhancement (min)	s.e.m.	No.
Normal s.n.	13	2.7	7
S.n. reinnervating skin	7	2.8	4
S.n. innervating muscle	67	8.3	4
Normal g.n.	64	7.5	13
G.n. reinnervating muscle	52	4.7	3
G.n. innervating skin	4	1.1	4

The duration of enhancement of flexion reflex (in min) induced by the injury discharge from axotomy of either the normal s.n. or g.n., or of self-anastomosed or of cross-anastomosed s.n. and g.n. innervating either appropriate or inappropriate targets.

Received 8 August; accepted 29 September 1989.

1. Barakat, I. & Droz, B. *Neuroscience* **28**, 39-47 (1989).
2. Frank, E. & Westerfield, M. *J. Physiol.* **324**, 495-505 (1982).
3. Marurich, M. F., Pourmehr, K. & Weston, J. A. *Devl. Biol.* **118**, 505-510 (1986).
4. Smith, C. L. & Frank, E. *Soc. Neurosci. Abs.* **10**, 100.34 (1984).
5. Smith, C. L. & Frank, E. *Soc. Neurosci. Abs.* **11**, 63 (1985).
6. Smith, C. L. & Frank, E. *J. comp. Neurol.* **269**, 96-108 (1988).
7. Smith, C. L. & Frank, E. *Brain Behav. Evol.* **31**, 227-242 (1988).
8. McMahon, S. B., Sykova, E., Wall, P. D., Woolf, C. J. & Gibson, S. *Neurosci. Lett.* **52**, 235-240 (1984).
9. McMahon, S. B. & Moore, C. E. *J. comp. Neurol.* **274**, 1-8 (1988).
10. Gibson, S. J. *et al. Peptides* **7**, 273-283 (1986).
11. Honda, C. N., Rethelyi, M. & Petrusz, P. *J. Neurosci.* **3**, 2183-2196 (1983).
12. Lewin, G. R. & McMahon, S. B. *Pflügers. Arch. ges. Physiol.* **411**, R179 (1988).
13. Lewin, G. R. & McMahon, S. B. *Eur. J. Neurosci. Sup.* **1**, 151.16 (1988).
14. McMahon, S. B. & Gibson, S. *Neurosci. Lett.* **73**, 9-15 (1987).
15. Wall, P. D. & Woolf, C. J. *J. Physiol.* **356**, 443-458 (1984).
16. Woolf, C. J. & Wall, P. D. *J. Neurosci.* **6**, 1433-1442 (1986).
17. Woolf, C. J. *Nature* **306**, 686-688 (1983).
18. Devor, M. & Wall, P. D. *Nature* **276**, 75-76 (1975).
19. Wall, P. D. & Devor, M. *Brain Res.* **209**, 95-111 (1981).
20. Wall, P. D., Fitzgerald, M. & Gibson, S. J. *Neurosci.* **6**, 2205-2215 (1981).
21. Merzenich, M. M. *et al. Neuroscience* **8**, 33-55 (1983).
22. Hylden, J. L. K., Nahin, R. L. & Dubner, R. *Brain Res.* **411**, 341-350 (1987).
23. Munson, J. B., Collins, W. F. & Mendell, L. M. in *The Current Status of Peripheral Nerve Regeneration* (Eds Gordon, T., Stein, R. B. & Smith, P. A.) 259-268 (Alan Liss, New York, 1988).
24. Lewin, G. R. & McMahon, S. B. *Soc. Neurosci. Abstr.* **15**, 13.9 (1989).

ACKNOWLEDGMENTS. We thank Caroline Abel for her invaluable technical help, and the Wellcome Trust and the Vivien L. Smith Foundation for financial support.

FMRFamide reverses protein phosphorylation produced by 5-HT and cAMP in *Aplysia* sensory neurons

J. David Sweatt*, Andrea Volterra, Brian Edmonds, Kevin A. Karl, Steven A. Siegelbaum & Eric R. Kandel

Howard Hughes Medical Institute and Center for Neurobiology and Behavior, College of Physicians and Surgeons of Columbia University, 722 West 168th Street, New York, New York 10032, USA

NEUROTRANSMITTERS can modulate neuronal activity through a variety of second messengers that act on ion channels and other substrate proteins¹⁻³. The most commonly described effector mechanism for second messengers in neurons depends on protein phosphorylation mediated by one of three sets of kinases: the cyclic AMP-dependent protein kinases^{4,5}, the Ca^{2+} -calmodulin-dependent protein kinases^{3,6}, and the Ca^{2+} -phospholipid-dependent protein kinases⁷. In addition, some neurotransmitters and second messengers can also inhibit protein phosphorylation by lowering cAMP levels (either by inhibiting adenylyl cyclase⁸ or activating phosphodiesterases⁹). This raises the question: can neurotransmitters also modulate neuronal activity by decreasing protein phosphorylation that is independent of cAMP? Various biochemical experiments show that a decrease in protein phosphorylation can arise through activation of a phosphatase^{10,11} or inhibition of kinases^{12,13}. In none of these cases, however, is the physiological role for the decrease in protein phosphorylation known. Here we report that in *Aplysia* sensory neurons, the presynaptic inhibitory transmitter FMRFamide decreases the resting levels of protein phosphorylation without altering the level of cAMP. Furthermore, FMRFamide overrides the cAMP-mediated enhancement of transmitter release produced by 5-hydroxytryptamine (5-HT), and concomitantly reverses the cAMP-dependent increase in protein phosphorylation produced by 5-HT. These findings indicate that a receptor-mediated decrease in protein phosphorylation may play an important part in the modulation of neurotransmitter release.

5-HT and FMRFamide produce opposing modulatory changes in the electrical excitability of the sensory neuron and in transmitter release from the sensory neuron presynaptic terminals¹⁴⁻¹⁶. By acting through cAMP-dependent phosphorylation, 5-HT produces prolonged all-or-none closure of the S-K⁺ channels, which results in a slow depolarization and an increase in the duration of the action potential, contributing to presynaptic facilitation¹⁴. By contrast, the neuropeptide FMRFamide, acting through the lipoxygenase metabolites of arachidonic acid¹⁷, produces an increase in the probability that an S-channel is open, which results in a slow hyperpolarization and a decrease in the action potential duration, which contributes to presynaptic inhibition¹⁵.

Moreover, the actions of 5-HT and FMRFamide are not symmetrical. Simultaneous application of the two transmitters does not result in simple algebraic summation of excitatory and inhibitory actions; rather, the inhibitory action of FMRFamide antagonizes or overrides the excitatory action of 5-HT. This override occurs at the level of single S-channel activity, where FMRFamide re-opens S-channels closed by 5-HT or cAMP¹⁵. Here we show that this override is also evident in the action of FMRFamide and 5-HT on transmitter release (Fig. 1a). The synaptic potential produced in a motoneuron in response to firing the sensory neuron was decreased by FMRFamide (presynaptic inhibition) to 11% of the control synaptic potential

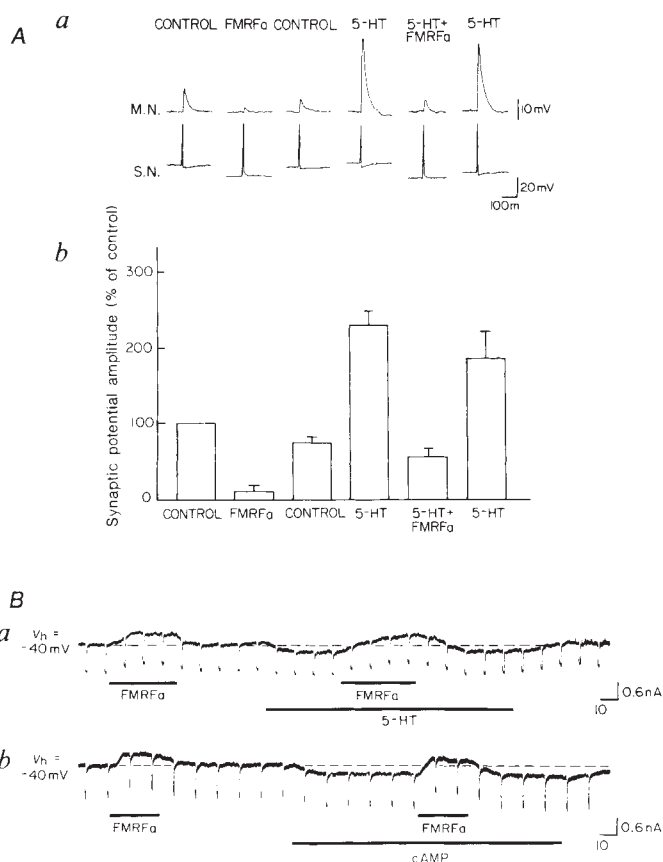


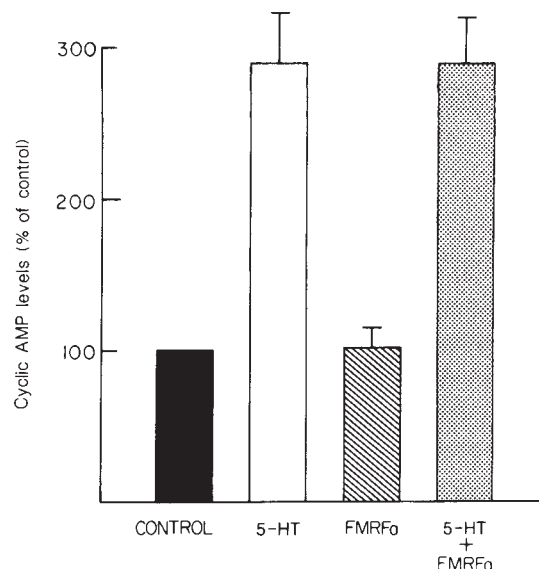
FIG. 1 FMRFamide (FMRFa) antagonizes the actions of 5-HT and cAMP. **A**, FMRFamide produces presynaptic inhibition and antagonizes presynaptic facilitation with 5-HT. **Aa**, Intracellular recordings of membrane potential from motor neurons (top) and sensory neurons (bottom) in response to brief depolarizing current stimuli to sensory neurons. Action potentials in sensory neurons elicit a fast excitatory postsynaptic potential (e.p.s.p.) in motor neurons. **Ab**, Average data from five experiments. Errors bars show s.e.m. Application of 10 μM FMRFamide alone inhibits the synaptic potential recorded in a follower neuron to 11% of that of the initial control. The same dose (10 μM) of 5-HT produces facilitation of the synaptic potential to 229% of that of the control. When FMRFamide is applied in the presence of 5-HT, the synaptic potential is reduced to a level slightly below that of the control (55%). **Ba**, Antagonism by FMRFamide of membrane current response to 5-HT. First application of FMRFamide (10 μM) produces a 450 pA increase in outward S-channel current ($V_m = -40$ mV). Application of 5-HT then produces a 270 pA decrease in outward S-channel current. Subsequent application of FMRFamide in the presence of 5-HT produces a 660 pA increase in outward current, so that net membrane current approaches peak outward level in FMRFamide alone. **Bb**, Antagonism of current response to cAMP by FMRFamide. V_h , Holding voltage. First application of FMRFamide produces a 420 pA increase in outward current. Next, application of 200 μM chlorophenyl-thio-cAMP causes a 300 pA decrease in outward current. Finally, application of FMRFamide in the presence of cAMP causes a 570 pA increase in outward current. On average, FMRFamide alone causes an increase in outward current of 0.28 ± 0.06 ($\pm$ s.d.; $n=6$; $V_m = -40$ mV), whereas cAMP alone causes a mean decrease in outward current of 0.2 ± 0.04 nA ($\pm$ s.d.; $n=6$; $V_m = -40$ mV). In the maintained presence of cAMP, FMRFamide induces a larger increase in outward current of 0.4 ± 0.007 nA ($\pm$ s.d.; $n=6$; $V_m = -40$ mV).

METHODS. **A**, Sensory neurons were grown in cell culture, with their follower motor neurons. Sensory neurons and motor neurons were impaled with a conventional intracellular microelectrode to record membrane potential under current clamp. Action potentials were elicited in the sensory neurons once every 40 s using a 5 ms depolarizing current pulse, and the fast e.p.s.p. were recorded in the motor neuron. **B**, Membrane current was measured from sensory neurons in intact abdominal ganglia by using a single microelectrode switched voltage-clamp (Axoclamp 2A, Axon Instruments). Electrode resistance was 10 M Ω and switching frequency was 6–8 kHz. Current records were filtered at 300 Hz. Hyperpolarizing command voltage-clamp steps of 10 mV were applied at 0.1 Hz.

* Present address: Division of Neuroscience, Baylor College of Medicine, 1 Baylor Plaza, Houston, Texas 77030, USA.

FIG. 2 Cyclic AMP levels in pleural sensory clusters after treatment with 5-HT or FMRFamide (FMRFa). Pleural sensory clusters were stimulated with 40 μ M 5-HT for 2 min in the presence of 100 μ M IBMX (1 min pre-incubation), or 10 μ M FMRFamide with 100 μ M IBMX for 1 min. Where 5-HT and FMRFamide were added together, 5-HT was added first for 1 min, then FMRFamide was added and the incubation continued for 1 min. Paired contra-lateral sensory clusters were used as controls and received no addition. Results shown are means $\pm$ s.e.m. for three separate experiments. Addition of IBMX alone cause no significant effect on cAMP levels when assayed at 3 min (control, 100%; IBMX, 79% $\pm$ 8%, n = 3). Previous studies have demonstrated that certain agonists decrease cAMP levels independent of adenylyl cyclase inhibition, at the level of phosphodiesterase activation⁹. The necessity of performing our cAMP assays on cells treated with phosphodiesterase inhibitor precluded us from this directly as a possible mechanism contributing to the inhibitory override.

METHODS. *Aplysia californica* (100–220 g) were obtained from the Marine Biological Laboratories mariculture facility at Woods Hole, Massachusetts. Animals were anaesthetized by injection of approximately one-half of their body weight of isotonic $MgCl_2$. Ring ganglia (cerebral, pleural and pedal ganglia) were removed and pleural sensory neuron clusters, with underlying neuropil, were dissected free (see ref. 19). These bilaterally symmetrical clusters, of $\sim$ 200 sensory neurons each, were kept as pairs from one animal. Each pair was used for a single experimental determination (assayed in triplicate), with one cluster serving as control and one as treated. This allowed us to perform intra-animal normalization and served to eliminate the inter-animal variability found for basal cAMP levels in *Aplysia* (see also ref. 19). The amount of protein in the left versus the right sensory cluster from one animal ($\sim$ 10 μ g) was found to be highly consistent. Thus, comparing left versus right is equivalent to normalizing each sample to protein. Cyclic AMP was assayed using a radioimmunoassay kit (Amersham). Incubations



were terminated by the addition of an ice-cold trichloroacetic acid solution to a final concentration of 10%. After 10 min on ice, samples were centrifuged, supernatants collected and ether-extracted, and cyclic AMP assayed. Results were calculated as the ratio of fmol cAMP for experimentals to fmol cAMP for controls. Control samples typically contained 50–75 fmol cAMP per sensory cell cluster.

(s.e.m. = 7.2, n = 5) and was increased by 5-HT (presynaptic facilitation) to 229% of the control (s.e.m. = 20, n = 5). When FMRFamide was applied in the presence of 5-HT, the facilitation of the synaptic potential was then overridden and the amplitude of the synaptic potential reduced to 55% of that of the control (s.e.m. = 12, n = 5).

As expected from the single channel results, the override is also evident in the total membrane current (Fig. 1b). Under voltage clamp, application of FMRFamide (10 μ M) alone produced an average increase in outward current of 0.23 ± 0.11 nA ($\pm$ s.d., n = 4; V_m = -40 mV). A mean decrease in outward current of 0.23 ± 0.12 nA (n = 4) was caused by 5-HT (10 μ M) alone. When FMRFamide was applied in the presence of 5-HT, however, the mean FMRFamide-induced current of 0.38 ± 0.10

(n = 4) was then greater than the FMRFamide-induced current in the absence of 5-HT. FMRFamide therefore not only reverses the decrease in outward current produced by 5-HT, but increases the net current to a level similar to the peak outward current reached with FMRFamide alone. FMRFamide exerts a similar override in the decrease in outward current produced by a membrane-permeable analogue of cAMP (Fig. 1b). Thus, at all three levels—transmitter release, membrane current and single channel activity—FMRFamide overrides the action of 5-HT. These results demonstrate that FMRFamide has at least two qualitatively distinct effects in sensory neurons; an inhibitory action when acting on its own, and an overriding action that opposes the action of 5-HT.

How, therefore, does FMRFamide exert its overriding action?

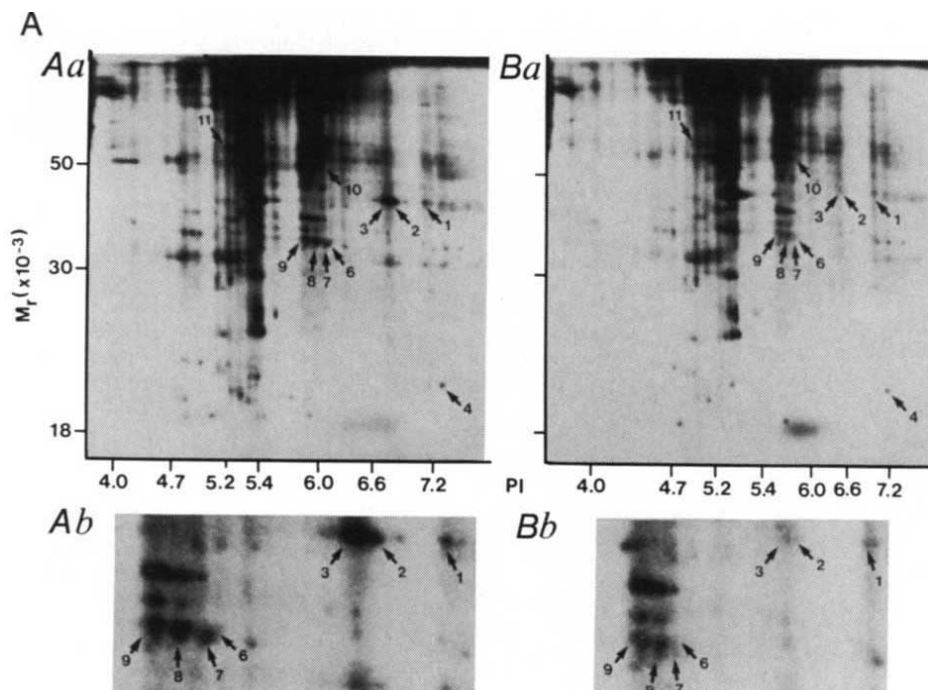


FIG. 3 A and B, Autoradiographs from control Aa and FMRFamide (Ba) treated cells. Autoradiographs are of 2-D gels run on 32 P-labelled sensory clusters. Labelling of sensory clusters, 2-D electrophoresis and autoradiography was performed as previously described²⁰. Ab and Bb, Expansion of the central regions of Aa and Bb. FMRFamide (10 μ M) was present for 1 min. In additional experiments (data not shown), no losses of protein spots in response to FMRFamide (10 μ M for 1 min) were observed in sensory neurons labelled for 24 h with [35 S]methionine (50 μ Ci per cluster), indicating that the loss of 32 P label was not due to a decrease in substrate protein mass.

Specifically, at which stage of the cAMP cascade does FMRFamide exert its antagonistic effect? Three sites of interaction seem possible. First, FMRFamide could override the cAMP cascade by inhibiting the adenylyl cyclase. Second, FMRFamide could stimulate phosphodiesterase activity thereby leading to an increase in the hydrolysis of cAMP (ref. 9). Third, FMRFamide, possibly acting through the metabolites of the lipoxygenase pathway¹⁷, could act at a later stage in the cAMP cascade to reverse phosphorylation of the substrate proteins either by inhibiting cAMP-dependent protein kinase or by activating a phosphoprotein phosphatase. To test these ideas, we studied the interaction of the two cascades at the biochemical level by examining cAMP production and protein phosphorylation.

Although adenylyl cyclase inhibition cannot *per se* explain the antagonism by FMRFamide of the effects of the analogue of cAMP (ref. 15 and Fig. 1b), inhibition of the cAMP cascade could still be a contributory mechanism in the antagonism by FMRFamide of the action of 5-HT. We therefore tested the ability of FMRFamide to inhibit adenylyl cyclase by using a radioimmunoassay for cAMP levels in the presence of the phosphodiesterase inhibitor 3-isobutyl-1-methyl-xanthine (IBMX) in *Aplysia* pleural sensory neuron clusters (Fig. 2). 5-HT increased the level of cAMP in *Aplysia* sensory neurons, whereas FMRFamide had no effect, thus confirming previous findings (refs 18 and 19, and T. W. Abrams *et al.*, personal communication). Moreover, FMRFamide did not attenuate the increase in cAMP produced by 5-HT (in agreement with K. A. Ocorr, M. Tabata and J. H. Byrne, personal communication).

These results indicate that the interaction between 5-HT and FMRFamide must occur after the rise in cAMP concentration, perhaps at the level of the phosphorylation and dephosphorylation of substrate proteins. We have previously characterized the effects of 5-HT and cAMP on substrate phosphorylation in the sensory neurons by using an intact-cell protein phosphorylation assay based on quantitative two-dimensional gel autoradiography²⁰. We have now used this assay to investigate the effects of FMRFamide on protein phosphorylation, and to search for possible sites of interaction between FMRFamide and the cAMP cascade.

We earlier found that, out of a total of 300–400 phosphoproteins in the sensory neurons resolved by two-dimensional gel electrophoresis²⁰, 5-HT and cAMP elicit an increase in phosphorylation of 17 proteins. By contrast, we found that FMRFamide elicits a decrease in protein phosphorylation (Figs 3 and 4). Thus, FMRFamide acting alone causes a decrease in the phosphorylation of a subset of 10 of the 17 proteins whose level of phosphorylation is increased by 5-HT or cAMP. No other phosphoproteins resolved on the gel were affected reproducibly. It is unlikely that the decrease in phosphorylation was caused by proteolysis, because autoradiographs from sensory neuron clusters labelled for 24 h with [³⁵S]methionine showed no significant loss of protein spots in response to FMRFamide.

Could a decrease in phosphorylation with FMRFamide contribute to the override of 5-HT's action? To address this question we examined the pattern of protein phosphorylation when FMRFamide was applied in the presence of 5-HT. We found that when FMRFamide is given in the presence of 5-HT it reverses the phosphorylation of all 17 proteins whose phosphorylation is increased in response to 5-HT (Fig. 4). Moreover, FMRFamide also decreased phosphorylation of proteins phosphorylated in response to application of a cAMP analogue. As the experiments with the cAMP analogue were done in the presence of the phosphodiesterase inhibitor IBMX, the decrease in phosphorylation cannot be explained by activation of a phosphodiesterase.

Recent evidence indicates that a single neurotransmitter can exert multiple physiological effects through separate biochemical pathways. Our previous experiments characterized two distinct actions of FMRFamide on S-channel function. Acting

on its own, FMRFamide leads to an increase in S-channel open probability (P_o), with no change in the number of active S-channels. When applied in the presence of cAMP or 5-HT, however, FMRFamide also reopens S-channels closed by these agents, leading to an increase in the number of open S-channels in addition to the increase P_o . Our present results indicate that these two actions may be mediated by distinct biochemical mechanisms. The increase in P_o with FMRFamide has been shown to be a direct action of 12-lipoxygenase metabolites of arachidonic acid on the S-channel, independent of protein phosphorylation or dephosphorylation (refs 21, 28). Here we show that FMRFamide also decreases the level of resting protein phosphorylation and reverses the increase in phosphorylation in response to 5-HT and cAMP, thereby providing a probable mechanism for the override of the physiological action of 5-HT and cAMP.

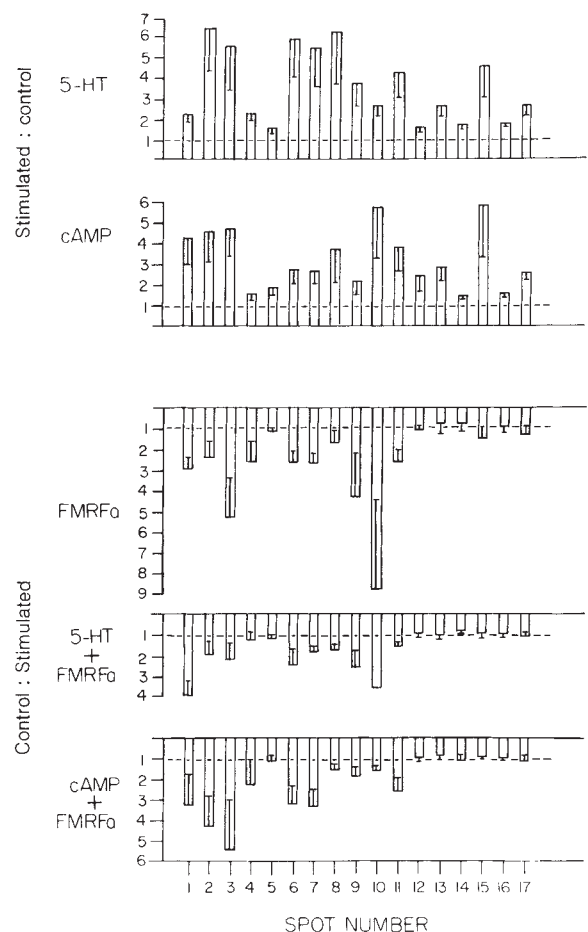


FIG. 4 Effects of various treatments on the phosphorylation of 17 selected proteins from *Aplysia* sensory neurons. Results shown are the means ($\pm$ s.e.m.) of the ratios of stimulated:control for 5-HT ($n=6$) and cAMP analogue ($n=7$), and control:stimulated for FMRFamide (FMRFa, $n=7$), FMRFa with 5-HT ($n=4$), and FMRFamide with cAMP analogue ($n=5$). Of ~300–400 phosphoprotein spots per gel, only these 17 were found to reproducibly change because of the indicated treatments. Quantitative information for the ³²P content of control and stimulated samples was determined by densitometric scanning, and spots were matched and ratios were calculated using the PDQuest computer analysis system²⁰. Treatments were 40 μ M 5-HT for 2 min, or 100 μ M 6-(4-chlorophenylthio)-cyclic AMP with 100 μ M phosphodiesterase inhibitor (either IBMX or R020-1724) for 10 min. Effects on protein phosphorylation were assayed at the end of the stimulus period. Where indicated, FMRFamide (10 μ M alone or in the presence of 5-HT or cAMP analogue with IBMX) was added 1 min before termination of the incubation. The results shown for 5-HT and cAMP analogue have been reported previously²⁰.

At present, we cannot distinguish whether this decrease in phosphorylation associated with the override in response to FMRFamide is due to kinase inhibition or phosphatase activation. Nor do our experiments by themselves prove a causal link between the decrease in phosphorylation and the override. Nonetheless, because the effects of 5-HT and cAMP on S-channel activity in the sensory neurons are mediated by cAMP-dependent phosphorylation, the decrease in phosphorylation with FMRFamide provides a plausible biochemical mechanism for the override.

Receptor-mediated decreases in protein phosphorylation (by inhibition of the kinase or activation of a phosphatase) may prove to be a widespread mechanism of cell modulation²². For

example, although the most common mechanism for antagonizing the cAMP cascade is through inhibition of adenylyl cyclase, some mechanisms for antagonism are known to operate in the face of elevated cAMP levels (for reviews, see refs 23–25). Thus, in addition to our previous findings that FMRFamide opens S-channels, there is evidence that in smooth muscle cells acetylcholine can override the effects of cAMP on M-current²⁶, and in pancreatic islet cells that α_2 -adrenergic agonists can inhibit the release of insulin induced by cAMP²⁷. These examples indicate that a receptor-mediated decrease in phosphorylation could serve in other cell systems to overcome the actions of the cAMP-dependent kinase, or even of other second messenger-dependent kinases. □

Received 7 August; accepted 18 September 1989.

- Kaczmarek, L. K. & Levitan, I. B. (eds) *Neuromodulation* (Oxford University Press, New York, 1987).
- Nestler, E. J. & Greengard, P. *Protein Phosphorylation in the Nervous System* (Wiley, New York, 1984).
- Siegelbaum, S. A. & Tsien, R. W. *Trends Neurosci.* **6**, 307–313 (1983).
- Levitan, I. J. *Membrane Biol.* **87**, 177–190 (1985).
- Kuo, J. F. & Greengard, P. *Proc. natn. Acad. Sci. U.S.A.* **64**, 1349–1355 (1969).
- Kennedy, M. B., Bennett, M. K. & Erond, N. E. *Proc. natn. Acad. Sci. U.S.A.* **80**, 7357–7361 (1983).
- Nishizuka, Y. *Science* **233**, 305–312 (1986).
- Gilman, A. G. *Cell* **36**, 577–579 (1984).
- Hughes, A. R., Martin, M. W. & Harden, T. K. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5680–5684 (1984).
- Stewart, A. A., Ingebritsen, T. S., Manalan, A., Klee, C. B. & Cohen, P. *FEBS Lett.* **137L**, 80–84 (1982).
- Hemmings, H. C., Greengard, P., Tung, H. Y. L. & Cohen, P. *Nature* **310**, 503–504 (1984).
- Walsh, D. A., Perkins, J. P. & Krebs, E. G. *J. biol. Chem.* **243**, 3763–3774 (1968).
- Walsh, D. A. *et al. J. biol. Chem.* **246**, 1977–1985 (1971).

- Siegelbaum, S. A., Camardo, J. S. & Kandel, E. R. *Nature* **299**, 413–417 (1982).
- Belardetti, F., Kandel, E. R. & Siegelbaum, S. A. *Nature* **325**, 153–156 (1987).
- Kandel, E. R. & Schwartz, J. H. *Science* **218**, 433–443 (1982).
- Piomelli, D. *et al. Nature* **328**, 38–43 (1987).
- Bernier, L., Castellucci, V. F., Kandel, E. R. & Schwartz, J. H. *J. Neurosci.* **2**, 1682–1691 (1983).
- Occorri, K. A. & Byrne, J. H. *Neurosci. Lett.* **55**, 113–118 (1985).
- Sweatt, J. D. & Kandel, E. R. *Nature* **339**, 51–54 (1989).
- Belardetti, F. *et al. Neuron* (in the press).
- Ingebritsen, T. S. & Cohen, P. *Science* **221**, 331–338 (1985).
- Limbird, L. E. & Sweatt, J. D. *In The Receptors* ed. P. M. Conn 281–301 (Academic, New York, 1985).
- Kato, R. & Nakaki, T. 1983. *Trends pharmacol. Sci.* **4**, 34–36 (1983).
- Limbird, L. E. *Fedn Proc.* **2**, 2686–2695 (1988).
- Simms, S. M., Singer, J. J. & Walsh, J. V. *Science* **239**, 190–193 (1988).
- Nakaki, T., Kakadate, T., Yamamoto, S. & Kato, R. *Life Sci.* **32**, 191–195 (1983).
- Buttner, N., Siegelbaum, S. A. & Volterra, A. *Nature* (in the press).

Inability of CD8 α' polypeptides to associate with p56^{lck} correlates with impaired function *in vitro* and lack of expression *in vivo*

Rose Zamoyska, Paula Derham, Scott D. Gorman^{†‡}, Paul von Hoegen,^{*} Joseph B. Bolen[‡], André Veillette[§] & Jane R. Parnes^{*}

Imperial Cancer Research Fund Tumour Immunology Unit, Biology Department, Medawar Building, University College London, Gower St., London WC1E 6BT, UK

^{*} Department of Medicine, Stanford University Medical Center, Stanford, California 94305, USA

[‡] Laboratory of Tumor Virus Biology, National Cancer Institute, Bethesda, Maryland 20892, USA

[§] McGill Cancer Center, McGill University, Montreal, Canada H3G 1Y6

T-CELL accessory molecules, particularly CD4 and CD8, seem to be involved in the control of T-cell activation by antigen. Precisely how such molecules operate is not fully understood, but evidence to date suggests a dual role, as receptors binding ligands on stimulator cells^{1,2} and by direct or indirect involvement in intracellular signalling events^{3–5}. In mouse, truncated 'tailless' CD8 molecules occur naturally (CD8 α' polypeptides) and although they are expressed on the surface of thymocytes, they are not expressed on the surface of mature T cells⁶. In this study, we show that truncated CD8 molecules are impaired in their ability to interact with the protein tyrosine kinase, p56^{lck}, and have decreased ability to restore immune responsiveness *in vitro*. Our data support a dual function for CD8 molecules correlated with expression of external domains and cytoplasmic domains, respectively. Both functions appear to be critical for a competent immune system *in vivo*.

We have previously shown that control of surface expression

of tailless CD8 α' polypeptides in mature peripheral T cells is at the level of protein transport⁷. That mature cells should be so stringent in excluding CD8 α' chain expression on the cell surface indicates that expression of these tailless molecules may have a negative influence. We tested this directly by transfecting individual complementary DNAs encoding CD8 α or α' polypeptides into a mouse T-cell hybridoma, DC27.10, which has been transfected with genes encoding the T-cell receptor (TCR) α and β chains specific for class I H-2K^b molecules. This hybridoma provides a convenient system in which to assay CD8 function as the transfectant is incapable of responding to antigen in the absence of CD8 molecules, although it will respond to the anti-TCR clonotype antibody, Désiré-1 (ref. 8). We selected variants of DC27.10 which lack expression of endogenous CD4 molecules before transfecting CD8 constructs. Both α and α' polypeptides can be expressed stably on the surface of this hybridoma in the absence of the CD8 β (Ly-3) product (Fig. 1a). The molecules are expressed as disulphide-bonded homodimers (data not shown). Several individual transfectants were analysed for their ability to respond to anti-clonotypic antibody bound to plastic (Fig. 1b) and antigen, namely K^b-transfected L cells (Fig. 1c). Expression of the truncated CD8 α' chain is sufficient to permit some restoration of the antigen response. Cells expressing the full length CD8 α polypeptide, however, respond much better. This is particularly apparent if you directly compare a pair of CD8 α and α' transfectants which express equivalent but lower levels of both transfected CD8 molecules and T-cell receptors. Transfectants a2 and b8, for example, express less CD8 (Fig. 1a) and respond less well to Désiré-1 antibody (Fig. 1b). The CD8 α transfectant, b8, responds well to antigen in contrast to the CD8 α' transfectant, a2, which is completely unresponsive. Comparison of transfectants expressing more CD8 and with higher responses to Désiré-1, a1 and b8.2 (Fig. 1a, b), shows that although the CD8 α' expressing cells can make some antigen response it is much weaker than that obtained from the CD8 α expressing cells (Fig. 1c). Indeed the antigen response of the brighter CD8 α' expressing transfectant is not even as good as that of the duller CD8 α expressing transfectant (compare a1 with b8). The fact that some restoration of response can occur in transfectants expressing CD8 α' chains indicates that the external domain alone does play a role in the antigen response.

[†] Present address; Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB1 2QP, UK.

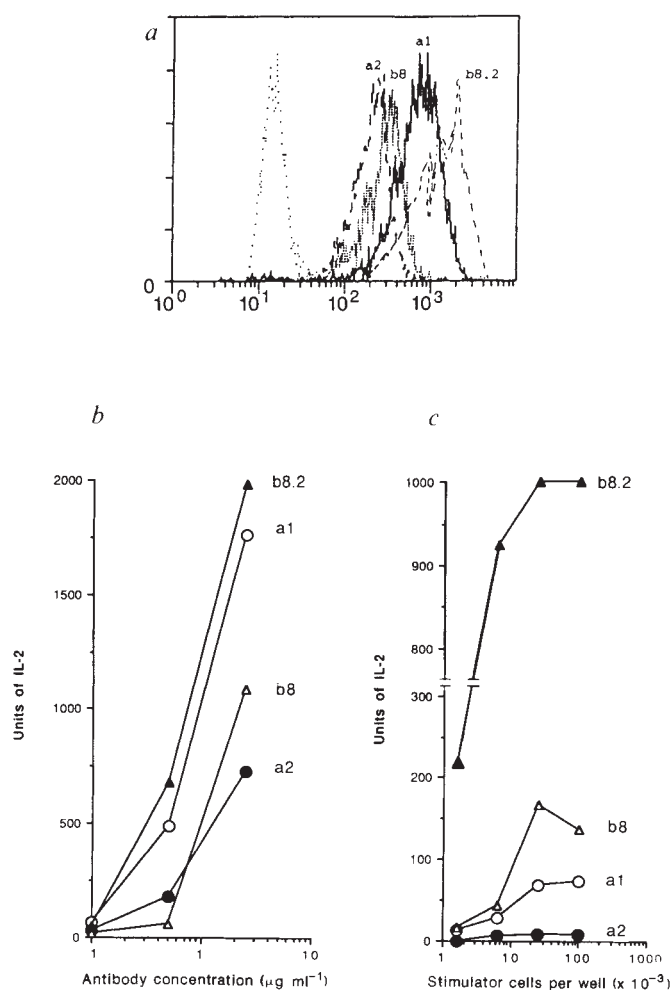


FIG. 1 Incremental restoration of antigen responsiveness by CD8 external and cytoplasmic domains. *a*, Cell-surface expression of CD8 homodimers by DC27.10 T cell hybridoma cells transfected with CD8 α (b8 and b8.2) and CD8 α' (a1 and a2) cDNA constructs and stained with anti-CD8 monoclonal antibody and fluoresceinisothiocyanate (FITC) labelled goat anti-rat second layer (Jackson Immunoresearch), or second layer alone (dotted profile). Fluorescence intensity, horizontal scale. The same transfectants expressing CD8 α (b8 and b8.2) and CD8 α' (a1 and a2), were assayed for IL-2 production when stimulated with anti-clonotypic antibody, Désiré-1, *b*, and with K^b-expressing transfected L cells, *c*.

METHODS. DC27.10 hybridoma cells were transfected by protoplast fusion²⁰ with the expression plasmid pH β APr-neo (ref. 21) containing CD8 α or α' full length cDNAs (ref. 22). Transfectants were selected with 600 $\mu\text{g ml}^{-1}$ (active concentration) G418, and expression of CD8 confirmed by staining with rat anti-mouse monoclonal antibody to Ly-2, YTS 169.4 (ref. 7), and FITC-labelled goat anti-rat (Jackson Immunoresearch). Transfectants were assayed for the ability to respond to anti-clonotypic antibody Désiré-1 by incubating purified antibody at the concentrations indicated in PBS medium on 96-well flat bottom tissue-culture plates for 30 min at 4 °C. Responder cells (1×10^5 per well) were added in tissue-culture media in a final volume of 200 μl and incubated at 37 °C. Antigen responses were assayed by incubating serial dilutions of K^b-transfected L cells with 1×10^5 responder cells in a final volume of 200 μl at 37 °C. For both anti-clonotypic and antigen responses supernatants were collected at 24 h and frozen at -20 °C until assayed for IL-2 content. IL-2 was assayed by titrating supernatants in triplicate in two-fold dilutions with 1×10^4 CTLL indicator cells per well. After 24 h, CTLL proliferation was assessed by pulsing with 20 μl of MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (Sigma), at 5 mg ml^{-1}) for 4 h and then solubilizing the cells overnight with 100 μl of 10% SDS in 0.01 M HCl. Plates were read on a VMAX kinetic plate reader at OD_{570–630 nm}. Total units of IL-2 per well were calculated by extrapolation of the linear portion of the graph of supernatant dilution versus optical density.

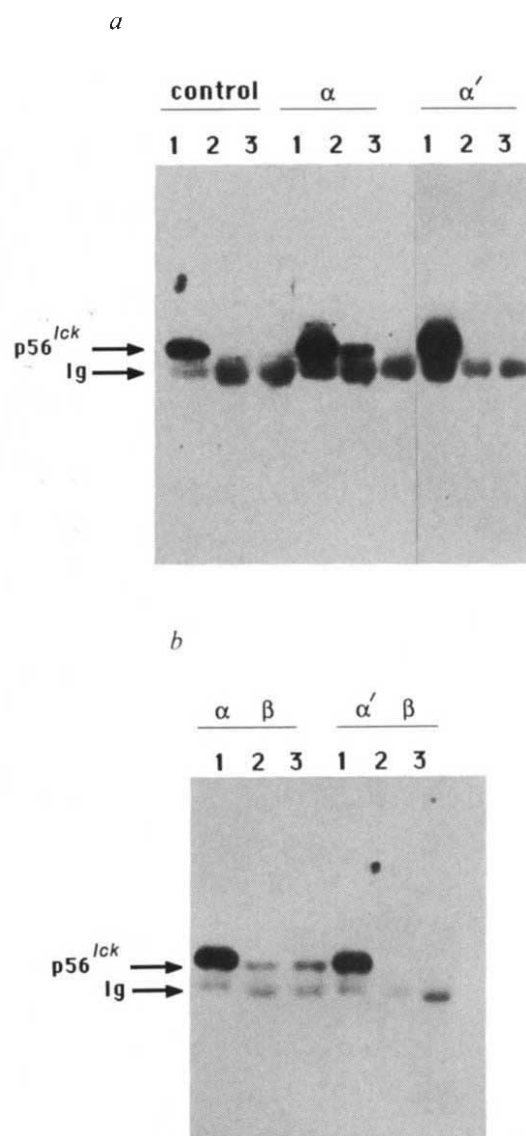


FIG. 2 p56^{lck} does not associate with CD8 α' or β polypeptides. *a*, Control, untransfected DC27.10, or DC27.10 transfected with CD8 α or α' cDNA, as indicated, were immunoprecipitated with: lane 1, anti-p56^{lck} antibody; lane 2, anti-CD8 α monoclonal antibody; lane 3, anti-CD8 β monoclonal antibody. The p56^{lck} content of the immunoprecipitates was analysed by immunoblotting as described¹¹, ~20% of total cellular p56^{lck} is associated with CD8 α , lane 2. Autoradiographic exposure: control and α panels, 12 h; α' panel, 24 h. *b*, Cells co-transfected with CD8 α and β cDNA, or with CD8 α' and β cDNA were immunoprecipitated and blotted as described above. Approximately 10% of total p56^{lck} is associated with CD8 $\alpha\beta$ heterodimers, lanes 2 and 3. Autoradiographic exposure 12 h.

METHODS. Cells were washed in cold PBS and lysed in TNE buffer (50 mM Tris (pH 8.0), 1% NP40 and 2 mM EDTA (pH 8.0)) containing protease inhibitors leupeptin and aprotinin (10 $\mu\text{g ml}^{-1}$) and the phosphotyrosyl phosphate inhibitor, sodium orthovanadate (100 μM). After clarification, immunoprecipitates were obtained from cellular lysates (150 μg) using specific anti-p56^{lck} antisera¹¹, anti-CD8 α monoclonal antibody (anti-Ly-2, 5.36; ref. 23), and anti-CD8 β monoclonal antibody (anti-Ly-3, 5.35; ref. 23).

This is presumably achieved by increasing the avidity of the TCR-antigen interaction through binding of CD8 to class I molecules, as has been demonstrated recently^{2,9}.

Both CD4 and CD8 molecules have been shown to co-precipitate with the protein tyrosine kinase p56^{lck} (refs 10,11). Also crosslinking with anti-CD4 antibodies results in rapid tyrosine phosphorylation of CD3 p21 polypeptides which has been ascribed to juxtaposition of CD4 and hence p56^{lck}, with CD3 (ref. 12). This has led to the suggestion that the participation

TABLE 1 Hybrid CD8/CD4 molecules respond to antigen only if they contain CD8 extracellular domains

Experiment	Transfectant	Accessory molecule			Response to: Désiré-1	K ^b L cells	K ^b spleen
		ext	tm	cyt			
1	a2	CD8	CD8	CD8 α'	730	7	—
	b8	CD8	CD8	CD8 α	1,085	137	—
	L1	CD8	CD8	CD4	193	87	—
2	a1	CD8	CD8	CD8 α'	298	62	—
	b2	CD8	CD8	CD8 α	130	265	—
	f6	CD8	CD4	CD4	200	189	—
3	b8.2	CD8	CD8	CD8 α	488	512	64
	c27.20	CD4	CD4	CD4	132	0	0
	e9	CD4	CD8	CD8 α	329	0	0
	k3	CD4	CD4	CD8 α	578	0	0
	j1	CD4	CD4	CD8 α'	276	0	0

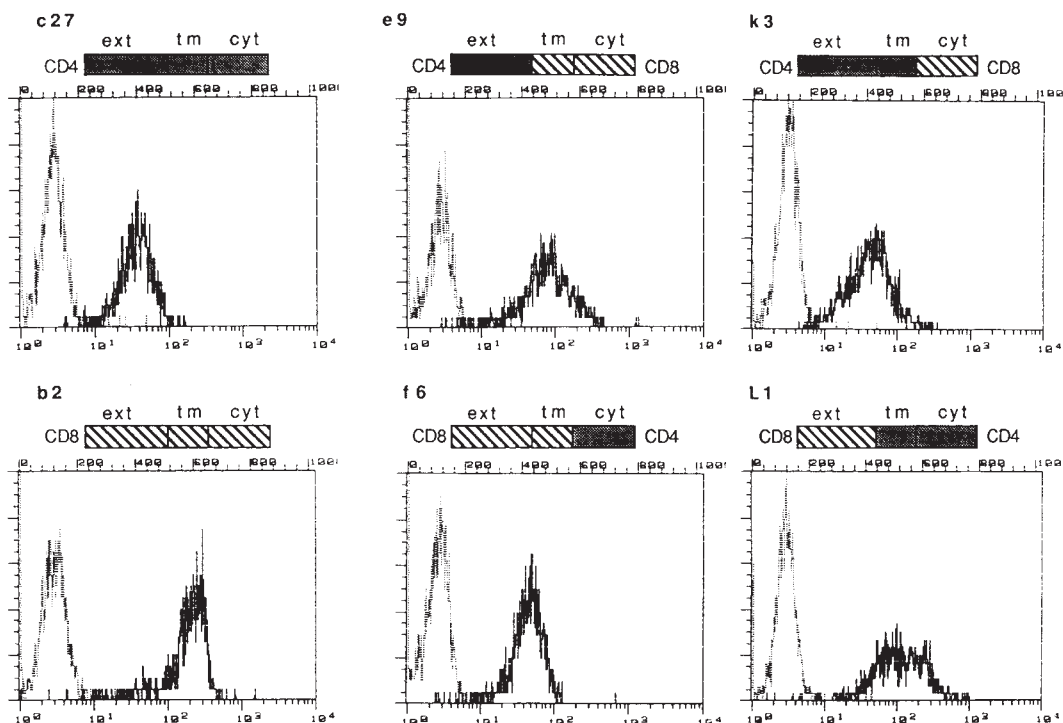
Transfectants expressing hybrid CD4/CD8 molecules as described in Fig. 3, were tested for response to anti-clonotypic antibody Désiré-1 (10 μ g/ml), 1×10^5 K^b expressing L cells per well or 1×10^6 C57B1/6 anti-thy-1 and complement treated spleen cells per well, as described in Fig. 1. Results are expressed as units of interleukin-2 (IL-2) extrapolated from titrating the supernatants on CTLL indicator cells (see Fig. 1).

of these accessory molecules in T-cell signalling occurs indirectly through association with this protein tyrosine kinase. One possible explanation for the decreased efficacy of the tailless CD8 α' chains would be an inability to associate with p56^{lck}, and thus participate in intracellular phosphorylation events. We tested this by immunoprecipitating CD8 α and α' transfectants with anti-p56^{lck} or anti-CD8 antibodies and immunoblotting to assess the presence of co-precipitating p56^{lck}. Figure 2a clearly demonstrates that in contrast to CD8 α chains, CD8 α' chains are not associated with p56^{lck}. The CD8 molecules on the surface of peripheral T cells are expressed normally as $\alpha\beta$ heterodimers, the products of two separate genes (Ly-2 and Ly-3 in mouse).

The CD8 β chain has a 19 amino-acid cytoplasmic tail and could potentially associate with p56^{lck}. We found, however, that this was not the case when we examined cells co-transfected with either CD8 α plus β or CD8 α' plus β complementary DNA. These transfectants show cell-surface expression of CD8 $\alpha\beta$ or $\alpha'\beta$ heterodimers respectively, but only the transfectants expressing CD8 $\alpha\beta$ show association with p56^{lck} (Fig. 2b). The CD8 $\alpha'\beta$ heterodimers are present intracellularly in mature murine T cells but are not expressed on the cell surface (R.Z., unpublished data). The demonstration that p56^{lck} cannot associate with this heterodimer indicates that expression of molecules capable of p56^{lck} association is critical for an optimally func-

FIG. 3 Expression of transfected CD4/CD8 hybrid molecules. DC27.10 hybridoma cells were transfected with hybrid cDNA molecules encoding assorted CD4 (filled boxes) and CD8 (dashed boxes) extracellular (ext) transmembrane (tm) and cytoplasmic (cyt) domains, as indicated over each FACS profile. Clones expressing CD4 extracellularly, c27, e9 and k3, were stained with FITC-labelled goat anti-rat antibodies alone (dashed profiles) and anti-CD4 (GK1.5) plus FITC-labelled goat anti-rat antibodies (solid profiles). Clones expressing CD8 extracellularly, b2, f6, and L1, were stained with FITC-labelled goat anti-rat antibodies alone (dashed profiles) and anti-CD8 (YTS169.4) plus FITC-labelled goat anti-rat antibodies (solid profiles). Fluorescence intensity, horizontal scale.

METHODS. Hybrid cDNA molecules were constructed by introducing complementary restriction enzyme sites (*Eco* RV at the extracellular/transmembrane junction, or a *Bst* XII site at the transmembrane/cytoplasmic junction) into CD8 and CD4 cDNAs using site-specific mutagenesis, and then swapping the domains between them. Oligonucleotides were designed to minimize amino-acid changes and to maintain the correct reading frame when introducing the restriction enzyme site, although in some cases conservative changes did result. No differences were observed in the functions of native and mutated but 'unshuffled' CD4 and CD8 molecules. The junctional amino-acid sequences for external/transmembrane domain shuffles were as follows:



native and mutated CD8-FACD/IYIW...; native CD4... VNQT/VFLA...; mutated CD4 (transfectant c27)... VNQT/ILA...; ext CD4/tm + cyt CD8 (transfectant e9)... NQTD/IYIW...; ext CD8/tm + cyt CD4 (transfectant f6)... FACD/ILA... Junctional amino-acid sequences for transmembrane/cytoplasmic domain shuffles were: native CD8-LICY/HRSR...; native CD4... LCCV/RCRH...; mutated CD8 (transfectant b2)... LICY/ARSR...; ext + tm CD4/cyt CD8 (transfectant k3)-LCCA/RSRL...; ext + tm CD8/cyt CD4 (transfectant L1)-LICY/RCRH... Hybrid cDNAs were subcloned into the β -actin expression vector and transfected as described in Fig. 1.

tional immune system *in vivo*. Given that the α chain alone is able to restore antigen responsiveness, it is not clear why there is a preference for expression CD8 $\alpha\beta$ heterodimers rather than CD8 $\alpha\alpha$ homodimers on mature T cells. It is possible that the CD8 β chain has a separate role during the development of $\alpha\beta$ TCR bearing cells in the thymus. This would correlate with the observation that intra-epithelial lymphocytes, which are mainly $\gamma\delta$ TCR bearing cells¹³, predominantly express CD8 α homodimers on their surface¹⁴.

The expression of CD4 or CD8 accessory molecules by T cells correlates strongly with whether T cells are restricted by class II or class I MHC molecules¹⁵. Studies in transgenic mice have shown that T cells expressing class I restricted TCR α and β transgenes favour the co-expression of CD8 molecules^{16,17}. Although this bias may be generated by co-selection of receptor and accessory molecules binding the same class I MHC structure, it is also possible that selection occurs based on preferential interactions between particular TCR and accessory molecule transmembrane or intra-cellular domains. We examined this possibility by constructing hybrid cDNA constructs between CD4 and CD8 genes which encode the extra-cellular portion of one together with either the cytoplasmic domain, or the transmembrane plus cytoplasmic domain of the other. It was previously shown that expression of CD4 molecules by DC27.10 is insufficient to restore the response to K^b even when stimulator cells express both class II and class I molecules⁸. If there is an interaction of the anti-K^b TCR with CD8 in the intra-cellular domains then the hybrids expressing CD4 externally and CD8 internally should be capable of responding to antigen.

Transfectants expressing all combinations of hybrid CD4/CD8 molecules were isolated and stained with appropriate monoclonal antibodies (Fig. 3). All of the hybrid constructs were expressed stably on the cell surface. Immunoprecipitation of these molecules (data not shown) revealed that all constructs expressing CD4 externally were monomers, in contrast to constructs expressing CD8 α externally, which were expressed as disulphide bonded homodimers. This indicates that interactions between CD8 external domains dictate that this molecule is expressed normally as a dimer. Transfectants which gave similar responses to anti-clonotypic antibodies were assayed for their ability to respond to antigen (Table 1). As K^b transfected L cells do not express class II MHC molecules, hybrid constructs expressing CD4 external domains were assayed on H-2K^b spleen cells. Despite the presence of the CD4 ligand on the stimulator cells, expression of these hybrid molecules was insufficient to restore antigen response. By contrast, transfectants expressing hybrid molecules containing CD8 α external domains, whether in combination with CD4 cytoplasmic domains or CD4 transmembrane and cytoplasmic domains, functioned extremely efficiently. The latter constructs were at least as effective as native CD8 α homodimers in conferring responsiveness to antigen. Thus, there does not appear to be preferential association of transmembrane or intracytoplasmic domains of CD8 with a class I specific cell receptor.

These data demonstrate an incremental restoration of full antigen responsiveness conferred by transfected CD8 molecules. The first step is the progress from complete absence to limited responsiveness conferred by high level expression of CD8 extracellular domains in combination with sufficient expression of TCRs. This property can be ascribed to increased avidity stabilizing the interaction between the T cell and the antigen presenting cell. It is interesting to note that in contrast to the experiments reported by Gay *et al.*¹⁸, our experimental system has an absolute requirement for both the TCR and accessory molecules to recognize class I MHC. The Désiré-1 TCR has low affinity for antigen¹⁹ and may only be stabilized by CD8 external domains binding to the same class I molecules as the TCR. This requirement may not be so obvious for higher affinity TCRs.

The second increment by which CD8 influences responsiveness relates to the presence of the cytoplasmic domain and

correlates with the ability to associate with p56^{lck}. This association has been demonstrated for both CD4 and CD8 molecules and can explain why CD4 intracellular domains are efficient substitutes for CD8 intracellular domains. These results are in contrast to hybrid CD8 molecules with class I cytoplasmic domains which are only as efficient as CD8 α' chains in restoring responsiveness (B. Malissen, personal communication). □

Received 5 June; accepted 25 September 1989.

- Doyle C. & Strominger, J. L. *Nature* **330**, 256 (1987).
- Norment, A. M., Salter, R. D., Parham, P., Engelhard, V. H. & Littman, D. R. *Nature* **336**, 79–81 (1988).
- Emmrich, F., Strittmatter, U. & Eichmann, K. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8298–8302 (1986).
- Ledbetter, J. A. *et al. Eur. J. Immun.* **18**, 525–532 (1988).
- Jönsson, J.-I., Boyce, N. W. & Eichmann, K. *Eur. J. Immun.* **19**, 253–260 (1989).
- Walker, I. D. *et al. Immunol. Rev.* **82**, 47–77 (1984).
- Zamoyska, R. & Parnes, J. R. *EMBO J.* **7**, 2359–2367 (1988).
- Gabert, J. *et al. Cell* **50**, 545–554 (1987).
- Rosenstein, Y., Ratnoffsky, S., Burakoff, S. J. & Herrmann, S. J. *J. exp. Med.* **169**, 149–160 (1989).
- Rudd, C. E., Trevelyan, J. M., Dasgupta, J. D., Wong, L. L. & Schlossman, S. F. *Proc. natn. Acad. Sci. U.S.A.* **85**, 5190–5196 (1988).
- Veillette, A., Bookman, M. A., Horak, E. & Bolan, J. B. *Cell* **55**, 301–308 (1988).
- Veillette, A., Bookman, M. A., Horak, E., Samelson, L. E. & Bolan, J. B. *Nature* **338**, 257–259 (1989).
- Goodman, T. & LeFrancis, L. *Nature* **33**, 855–858 (1988).
- Parrott, D. M. V. *et al. Ann. N.Y. Acad. Sci.* **409**, 307–320 (1983).
- Swain, S. L. *Immunol. Rev.* **74**, 129–142 (1983).
- Kisielow, P., Teh, H. S., Blüthmann, H. & von Boehmer, H. *Nature* **335**, 730–733 (1988).
- Sha, W. C. *et al. Nature* **336**, 73–76 (1988).
- Gay, D. *et al. Nature* **328**, 626–629 (1987).
- Hua, C., Boyer, C., Guimenzanes, A., Alpert, F. & Schmitt-Verhulst, A. M. *J. Immun.* **136**, 1927–1936 (1986).
- Oi, V. T., Morrison, S. L., Herzenberg, L. A. & Berg, P. *Proc. natn. Acad. Sci. U.S.A.* **80**, 825–829 (1983).
- Gunning, P., Leavitt, J., Muscat, G., Ng, S.-Y. & Keddes, L. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4831–4835 (1987).
- Zamoyska, R., Vollmer, A. C., Sizer, K. C., Liaw, C. W. & Parnes, J. R. *Cell* **43**, 153–163 (1985).
- Ledbetter, J. A. & Herzenberg, L. A. *Immunol. Rev.* **47**, 63–90 (1979).

ACKNOWLEDGEMENTS. We thank Hans Stauss for K^b transfected L cells and discussion, Dominique Blanc, Jean Gabert and Bernard Malissen for the DC27.10 cell line and discussing their unpublished data, and Paul Travers, Lucia Casabo, Sarah Bell and Owen Williams for reading of the manuscript. S.D.G. was a recipient of a NCI fellowship, P.V.H. is supported by a postdoctoral fellowship from the Deutsche Forschungsgemeinschaft, A.V. is supported by the National Cancer Institute of Canada. J.R.P. holds an Established Investigator Award from the American Heart Association and is supported by grants from the NIH.

Genetic imprinting suggested by maternal heterodisomy in non-deletion Prader-Willi syndrome

Robert D. Nicholls*†‡, Joan H. M. Knoll†, Merlin G. Butler§, Susan Karam|| & Marc Lalonde*†¶

* Howard Hughes Medical Institute, and † Genetics Division and Mental Retardation Center, The Children's Hospital, Department of Pediatrics, Harvard Medical School, 300 Longwood Avenue, Boston, Massachusetts 02115, USA

§ Division of Genetics, Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA

|| Michigan State University Medical School, East Lansing, Michigan 48824, USA

PRADER-WILLI syndrome (PWS) is the most common form of dysmorphic genetic obesity associated with mental retardation^{1,2}. About 60% of cases have a cytological deletion of chromosome 15q11q13 (refs 2, 3). These deletions occur *de novo* exclusively on the paternal chromosome^{4,5}. By contrast, Angelman syndrome (AS) is a very different clinical disorder and is also associated with deletions of region 15q11q13 (refs 6–8), indistinguishable from those in PWS^{6,8} except that they occur *de novo* on the maternal chromosome⁶. The parental origin of the affected chromosomes 15 in these disorders could, therefore, be a contributory factor in determining their clinical phenotypes. We have now used

‡ Present address: Department of Neuroscience, University of Florida College of Medicine, Gainesville, Florida 32610, USA.

¶ To whom correspondence should be addressed.

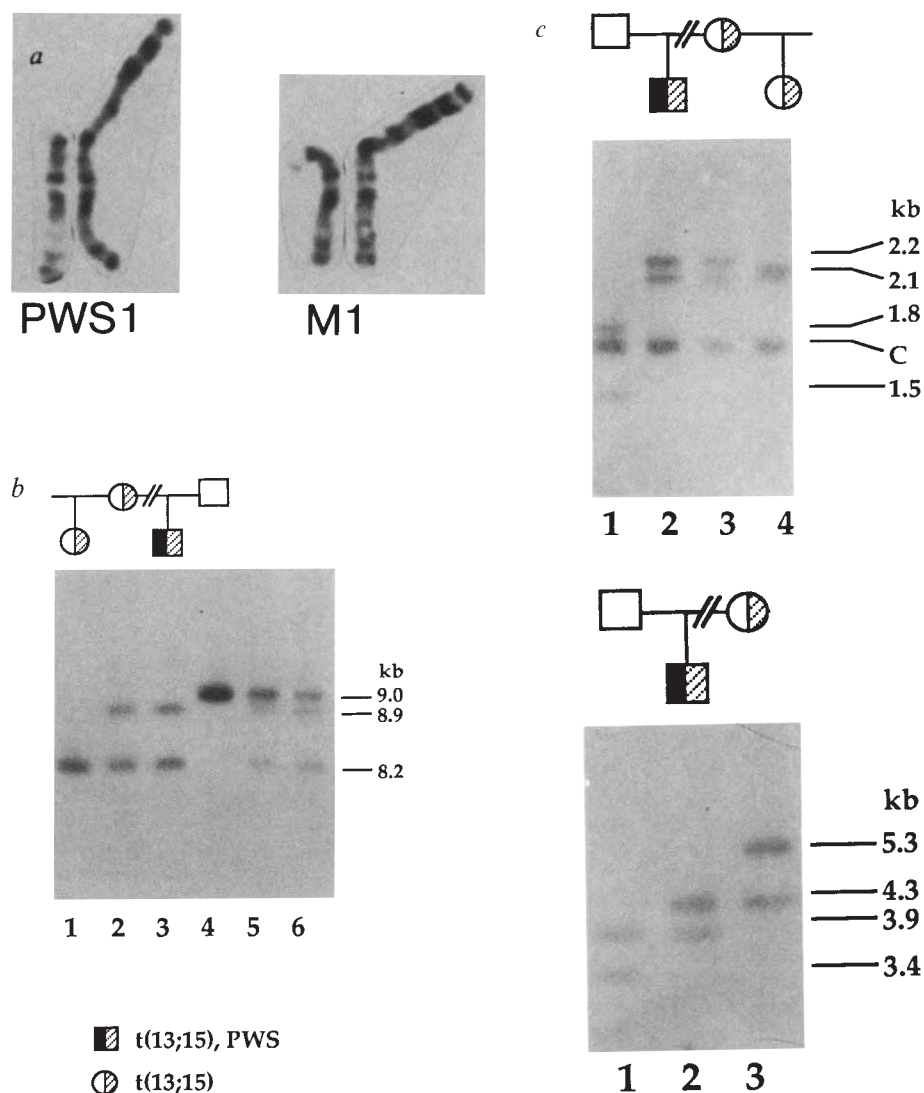
cloned DNA markers specific for the 15q11q13 subregion^{5,9,10} to determine the parental origin of chromosome 15 in PWS individuals not having cytogenetic deletions; these individuals account for almost all of the remaining 40% of PWS cases. Proband in two families displayed maternal uniparental disomy for chromosome 15q11q13. This is the first demonstration that maternal heterodisomy—the presence of two different chromosome 15s derived from the mother—can be associated with a human genetic disease. The absence of a paternal contribution of genes in region 15q11q13, as found in PWS deletion cases^{4,5}, rather than a mutation in a specific gene(s) in this region may result in expression of the clinical phenotype. Thus, we conclude that a gene or genes in region 15q11q13 must be inherited from each parent for normal human development.

Given the unusual parental origins of deletion chromosomes associated with PWS and AS, we attempted to determine the parental origins of chromosomes 15 in non-deletion cases of PWS. For each family, we performed restriction fragment length polymorphism (RFLP) analyses with nine RFLPs at seven loci specific for proximal chromosome 15q (refs 5, 11). The first family that we studied has a proband with a phenotype typical of PWS (refs 1, 2), that is, hypotonia and growth delay in infancy, short stature, hyperphagia with consequent obesity, small hands and feet, hypogonadism and mild mental retardation. This family also displays inheritance of a balanced robertsonian

translocation, t(13; 15) (Fig. 1a). Association of this translocation with the phenotype of PWS is highly unlikely because unaffected maternal relatives carry the same balanced translocation (Fig. 1a), and the translocation breakpoint is at the centromere of chromosome 15, some distance from the critical region at 15q11q13. In this family the PWS proband (PWS1) had inherited two maternal alleles but no paternal allele (Fig. 1b) specific for probe 3-21, which maps to the region absent in PWS patients with deletions^{5,9,10}. The 8.2-kilobase (kb) *TaqI* fragment detected by probe 3-21 (Fig. 1b) is shared by the three individuals (PWS1, his mother M1, and half-sister) carrying the t(13; 15) translocation and is, therefore, a marker for this chromosome. The 8.9-kb *TaqI* fragment which hybridized to probe 3-21 is present in both PWS1 and his mother and is not present in the other individuals of the family, clearly indicating the maternal derivation of the non-translocated chromosome 15 in PWS1, at least for region 15q11q13 surrounding the probe 3-21 locus. We confirmed and extended this result by using CMW-1, a multiallelic DNA marker¹¹ that maps distal to the region of deletions associated with PWS (unpublished result). The segregation pattern again showed two maternal alleles but no paternal allele for the proband (Fig. 1c). We excluded the possibility of non-paternity by analysis with a cloned fragment (3'HVR) from the 3' hypervariable region of the α -globin locus on chromosome 16 (ref. 12, Fig. 1d).

FIG. 1 Maternal uniparental disomy in a PWS patient with a balanced Robertsonian translocation. **a**, Normal chromosome 15 and translocated homologue from patient PWS1 (left) and his mother M1 (right). Other unaffected maternal relatives, including the half-sister of PWS1 carry the same balanced chromosome 13q15q robertsonian translocation. **b**, Maternal origin of the two chromosomes 15 in patient PWS1. An 8.2-kb *TaqI* allele of probe 3-21 (D15S10) is shared by PWS1 (lane 3), his mother M1 (lane 2) and half-sister (lane 1), whereas an 8.9-kb band is shared by PWS1 and M1. PWS1 does not inherit a 9.0-kb paternal allele (F1, lane 4). Mixing experiments (lane 5, PWS1 and F1; lane 6, M1 and F1) show the identity of the three alleles in this new RFLP system. **c**, Presence of two maternal alleles only for CMW-1 (D15S24) in PWS1. PWS1 (lane 2) shares two *TaqI* restriction fragments, of 2.1 and 2.2 kb, with M1 (lane 3) but has neither of the 1.5- or 1.8-kb paternal alleles (F1, lane 1). The half-sister (lane 4) displays the maternal 2.1-kb and a 2.15-kb allele. The band at 1.7 kb is constant, C. **d**, Exclusion of non-paternity for PWS1. *PvuII*-digested DNA was hybridized to the α -globin 3'HVR marker (D16S85), one of the most highly polymorphic markers in the human genome¹². PWS1 (lane 2) shows mendelian inheritance of one paternal 3.9-kb allele (lane 1) and one maternal 4.3-kb allele (lane 3). The half-sister inherits her mother's 5.3-kb allele (result not shown).

METHODS. DNA was isolated from blood and/or cell lines from each individual as described previously⁵. Hybridizations were carried out as previously described⁵. High-resolution chromosome analyses from peripheral blood lymphocytes²⁹ were performed by Giemsa-trypsin banding³⁰.



The proband (PWS2) of the second family also has classical PWS (refs 13, 14). Cytogenetically, PWS2 has two intact chromosome 15s (Fig. 2a). We demonstrated the lack of paternal alleles for two loci by haplotype analysis (Fig. 2b, c). With probe 34, the father is homozygous for 6.5 kb alleles, whereas the mother is homozygous for 6.3 kb alleles (Fig. 2b). For the proband, we found only a 6.3-kb band (Fig. 2b), consistent with the maternal origin of both cytogenetically normal chromosomes 15. Likewise, only maternal alleles specific for probe IR39d had been inherited (Fig. 2c). By dosage analyses we demonstrated that PWS2 has two copies of each of these maternally derived alleles (Fig. 2d and results not shown), indicating that loci detected

by probes 34 and IR39d had not been deleted. These two loci are flanked by loci detected by probe IR4-3R (results not shown) and probe IR10-1, respectively, (Fig. 2b) that are heterozygous in the proband. To rule out the possibility of a submicroscopic deletion, we analysed the long-range structure of the 15q11q13 region by pulsed-field gel electrophoresis (PFGE). This showed that probe IR4-3R and probe 34 detect the same large unaltered 2,500-kb *NorI* fragment in PWS2 (Fig. 3), which provides strong evidence that the absence of a paternal allele specific for probe 34 does not represent a microdeletion. Furthermore, we detected a band corresponding to a fragment of the same size (2,500 kb) in the mother, despite differential methylation leading to partial

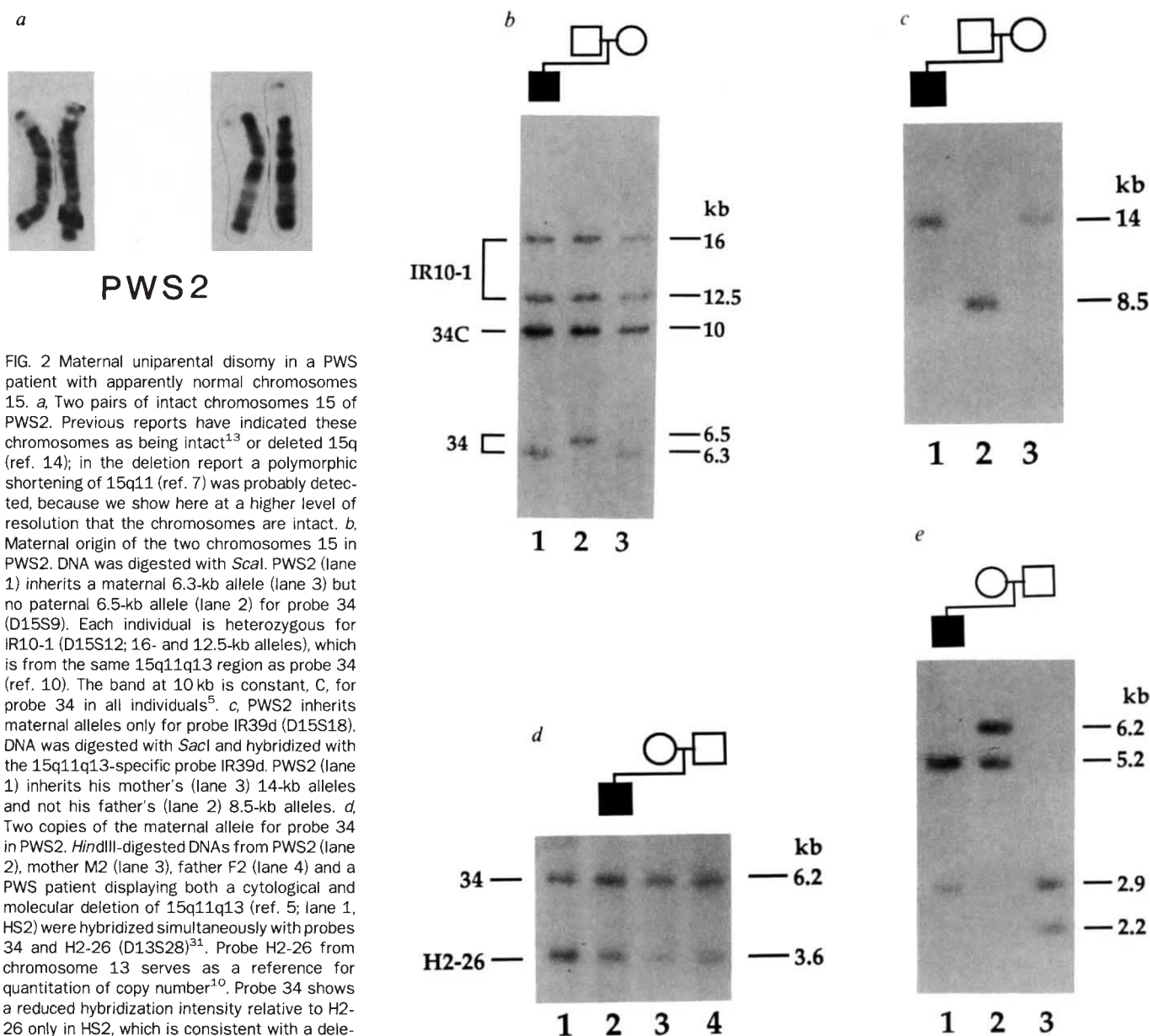


FIG. 2 Maternal uniparental disomy in a PWS patient with apparently normal chromosomes 15. a, Two pairs of intact chromosomes 15 of PWS2. Previous reports have indicated these chromosomes as being intact¹³ or deleted 15q (ref. 14); in the deletion report a polymorphic shortening of 15q11 (ref. 7) was probably detected, because we show here at a higher level of resolution that the chromosomes are intact. b, Maternal origin of the two chromosomes 15 in PWS2. DNA was digested with *ScaI*. PWS2 (lane 1) inherits a maternal 6.3-kb allele (lane 3) but no paternal 6.5-kb allele (lane 2) for probe 34 (D15S9). Each individual is heterozygous for IR10-1 (D15S12; 16- and 12.5-kb alleles), which is from the same 15q11q13 region as probe 34 (ref. 10). The band at 10 kb is constant, C, for probe 34 in all individuals⁵. c, PWS2 inherits maternal alleles only for probe IR39d (D15S18). DNA was digested with *SacI* and hybridized with the 15q11q13-specific probe IR39d. PWS2 (lane 1) inherits his mother's (lane 3) 14-kb alleles and not his father's (lane 2) 8.5-kb alleles. d, Two copies of the maternal allele for probe 34 in PWS2. *HindIII*-digested DNAs from PWS2 (lane 2), mother M2 (lane 3), father F2 (lane 4) and a PWS patient displaying both a cytological and molecular deletion of 15q11q13 (ref. 5; lane 1, HS2) were hybridized simultaneously with probes 34 and H2-26 (D13S28)³¹. Probe H2-26 from chromosome 13 serves as a reference for quantitation of copy number¹⁰. Probe 34 shows a reduced hybridization intensity relative to H2-26 only in HS2, which is consistent with a deletion of 15q11q13 in this individual. No reduction in probe 34 hybridization intensity is observed in PWS2 or his parents. To quantitate the intensity of the hybridization bands, the autoradiogram was scanned with an LKB Ultrascan XL laser densitometer (Pharmacia LKB Biotechnology). The area ratio of the probe 34 to probe H2-26 peaks was calculated for each lane. The number of copies of probe 34 per genome was obtained by setting the ratio in lane 3 (M2, mother) to 2 and normalizing the values in the other lanes to this value. The number of copies of probe 34 per genome determined by this method were 1.0 in lane 1 (the HS2

deletion), 1.8 in lane 2 (PWS2) and 2.2 in lane 4 (F2, father). e, Exclusion of non-paternity for PWS2. One maternal 5.2-kb (lane 2) and one paternal 2.9-kb (lane 3) *PvuII* allele are inherited by PWS2 (lane 1) for the 3'HVR. METHODS. DNA and cytogenetic studies were as for Fig. 1. Probe IR39d is a subclone of IR39 (ref. 10), which lacks repetitive sequences. PWS2 has been described before as patient number 5 (ref. 14) and patient number 31 (ref. 13).

NotI digestion in this region of the genome (Fig. 3) and in other normal individuals (results not shown). We again excluded the possibility of non-paternity using the 3'HVR fragment, which detected an entirely different set of alleles in the PWS2 family (Fig. 2e) from those that it detected in the PWS1 family (Fig. 1d).

The results presented here indicate that both PWS probands inherited two different, intact chromosome 15q11q13 regions from their mothers. Maternal heterodisomy—a newly defined form of uniparental disomy^{15,16}—for at least the critical region of chromosome 15, could thus have an aetiological role in PWS. The extent of disomy remains to be determined by genetic analysis with other markers. The disomy extends beyond the PWS critical region identified from deletion studies because locus CMW-1, which maps distal to the deletion region, is

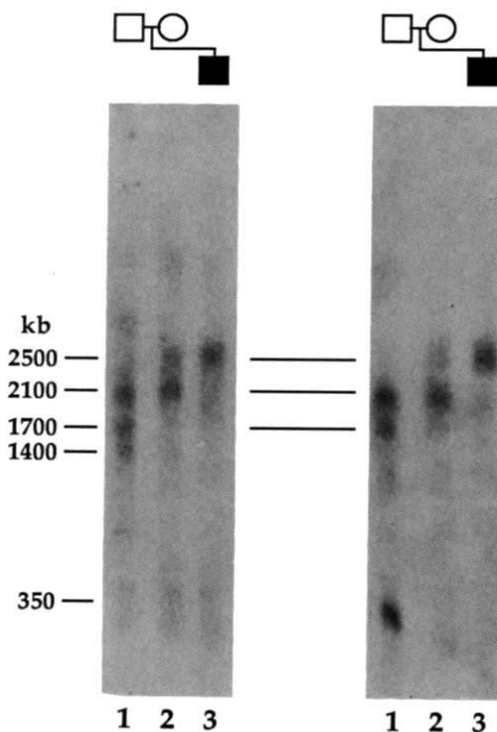


FIG. 3 PFGE provides evidence for the lack of a submicroscopic deletion in PWS patients with uniparental disomy. *NotI* fragments are shown for probes IR4-3R (D15S11) (left) and 34 (right) in PWS2 (lane 3), M2 (lane 2) and F2 (lane 1). Multiple bands are indicative of partial cleavage at restriction sites when CpG is methylated³² and a similar pattern of multiple bands has been observed in other normal controls (results not shown). PWS2 shares a normal 2,500-kb band with his mother, although in the latter this results from partial digestion at a *NotI* site in this region of the genome. The same 2,500-kb band has been detected in other normal individuals. The larger three fragments are common to probes IR4-3R, 189-1 (results not shown) and 34, showing that these loci are closely linked in region 15q11q13. These probes and a 1,300-kb fragment detected with probes 3-21 and IR10-1 (unpublished data) are deleted in all PWS¹⁰ and AS⁶ patients with cytological deletions. Additional PFGE analysis (results not shown) also indicates that PWS2 displays PFGE restriction fragments of normal size for IR10-1 in a *NotI* digest (1,300 kb) and for IR39d in a *BssHII* digest (250 kb). PWS1 also seems to be intact for the probe 34 and 3-21 loci because the restriction fragments detected using *NotI* and *BssHII* are unaltered compared with those of normal individuals.

METHODS. PFGE analyses were performed as described previously³³ using a modification of techniques for DNA preparation and digestion in agarose blocks³⁴. Electrophoresis was performed in 0.8% agarose in TBE (89 mM Tris buffer, pH 8.0, 89 mM boric acid, 2 mM EDTA) at 22 °C and was divided into three 48-h intervals during which the forward polarity pulse durations were varied from 75–600 s, 600–2,500 s and 2,500–3,000 s and field intensities were set at +2, +1.7 and +1.3 V cm⁻¹, respectively. The ratios of the duration and intensity of the inverse polarity pulse relative to the forward were 0.4 and 0.5, respectively.

disomic in PWS1. Several predictions can be made about an association between uniparental disomy and the aetiology of PWS. First, there should be no chromosome deletion, as is shown in this study by using both cytogenetic and molecular genetic techniques. Additional PFGE mapping data for PWS1 and PWS2 (Fig. 3 legend) confirmed that a large part (4–5 megabases) of the 15q11q13 region is intact in these two patients. Second, there should be no specific gene mutation. Although point mutations cannot be ruled out until the gene(s) responsible for PWS are isolated, these seem unlikely because each case displays two independent maternal contributions to region 15q11q13 and the disorder is genetically dominant, which would rule out mechanisms such as the uncovering of a recessive mutation. Finally, the frequency of maternal disomy should be high in PWS patients with normal chromosomes. Consistent with this prediction, our preliminary data on four additional families strongly indicates that all of the PWS patients show maternal uniparental disomy.

It seems that the clinical phenotype of the two PWS probands arises from the absence of a paternal contribution to region 15q11q13, rather than from a specific gene mutation. This implies functional differences in alleles of a gene or genes from this region of the genome that depend on the sex of the transmitting parent (genetic imprinting^{18–21}). Under this scheme, normal human development would require genetic input from both parents, whereas the absence of a paternal contribution to region 15q11q13, whether by paternal deletion^{4,5} or maternal uniparental disomy (as demonstrated here), would result in PWS. Consistent with this hypothesis is our recent finding of differential transmission of parental alleles in PWS and AS (ref. 6). It is conceivable, therefore, that the absence of a maternal contribution to the same 15q11q13 region could result in a different disorder, AS.

Experimental evidence for genetic imprinting has been largely restricted to the mouse in which a requirement for a contribution of both the maternal and paternal genome for normal development has been demonstrated^{18–22}. Transgenic studies have provided preliminary evidence that the chromatin alteration involved in the differential modification and expression of parental alleles is DNA methylation^{18–22}, a process also associated with transcriptional regulation of gene activity²³. In humans, evidence of a developmental requirement for both parental genomes is provided by the finding that only the paternal genome is present in complete hydatidiform moles²⁴. Furthermore, differential transmission of parental alleles in several disorders such as Huntington's chorea^{18,25}, and somatic changes in childhood tumours such as Wilms tumour, osteosarcoma (reviewed in ref. 26) and rhabdomyosarcoma²⁷ could involve genome imprinting. Growth failure in two individuals with isodisomy¹⁵ for chromosome 7 (refs 16, 28) could indicate genetic imprinting¹⁶, although the aetiology of cystic fibrosis (CF) in these patients is homozygosity for a mutant CF allele inherited from the mother^{16,28}.

In conclusion, the association of PWS with maternal uniparental disomy for region 15q11q13 implicates a role for genetic imprinting in the aetiology of the PWS phenotype. Similar phenomena could contribute to other clinical syndromes. Isolation of the gene(s) responsible for PWS and other disorders in which genetic imprinting plays a part might be possible by identifying a gene that shows a differential modification of its parental alleles. Our findings for PWS represent a step towards understanding the developmental basis of this common human genetic disorder. □

Received 31 July; accepted 13 October 1989.

1. Cassidy, S. B. *Curr. Problems Pediatr.* **14**, 1–55 (1984).
2. Butler, M. G. *Am. J. med. Genet.* (in the press).
3. Ledbetter, D. H., Greenberg, F., Holm, V. A. & Cassidy, S. B. *Am. J. med. Genet.* **28**, 779–790 (1987).
4. Butler, M. G., Meaney, F. J. & Palmer, C. G. *Am. J. med. Genet.* **23**, 793–809 (1986).
5. Nicholls, R. D. et al. *Am. J. med. Genet.* **33**, 66–77 (1989).
6. Knoll, J. H. M. et al. *Am. J. med. Genet.* **32**, 285–290 (1989).

7. Pembrey, M. et al. *J. Med. Genet.* **26**, 73-77 (1989).
8. Donlon, T. A. *Hum. Genet.* **80**, 322-328 (1988).
9. Donlon, T. A., Lalande, M., Wyman, A., Bruns, G. & Latt, S. A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4408-4412 (1986).
10. Tantravahi, U. et al. *Am. J. med. Genet.* **33**, 78-87 (1989).
11. Rich, D. C., Witkowski, C. M., Summers, K. M., van Tuinen, P. & Ledbetter, D. H. *Nucleic Acids Res.* **16**, 8740 (1988).
12. Higgs, D. R. et al. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5165-5169 (1986).
13. Butler, M. G. & Meaney, F. J. *Am. J. med. Genet.* **26**, 445-455 (1987).
14. Butler, M. G. *Am. J. hum. Genet.* **45**, 140-146 (1989).
15. Engel, E. *Am. J. med. Genet.* **6**, 137-143 (1980).
16. Spence, J. E. et al. *Am. J. hum. Genet.* **42**, 217-226 (1988).
17. Warburton, D. *Am. J. hum. Genet.* **42**, 215-216 (1988).
18. Solter, D. A. *Rev. Genet.* **22**, 127-146 (1988).
19. Sapienza, C. *Annis N.Y. Acad. Sci.* (in the press).
20. Monk, M. *Genes Dev.* **2**, 921-925 (1988).
21. Holliday, R. *Sci. Am.* **260**, 60-73 (1989).
22. Surani, M. A., Reik, W. & Allen, N. D. *Trends Genet.* **4**, 59-62 (1988).
23. Cedar, H. *Cell* **53**, 3-4 (1988).
24. Kajii, T. & Ohama, K. *Nature* **268**, 633-634 (1977).
25. Reik, W. *J. med. Genet.* **25**, 805-808 (1988).
26. Reik, W. & Surani, M. A. *Nature* **338**, 112-113 (1989).
27. Fischel-Ghodsian, N., Nicholls, R. D. & Higgs, D. R. *Nucleic Acids Res.* **15**, 6197-6207 (1987).
28. Voss, R. et al. *Am. J. hum. Genet.* **45**, 373-380 (1989).
29. Yunis, J. J. *Science* **191**, 1268-1270 (1976).
30. Seabright, M. *Lancet* **ii**, 971-972 (1971).
31. Lalande, M. et al. *Cancer Genet. Cytogenet.* **13**, 283-295 (1984).
32. Fischel-Ghodsian, N., Nicholls, R. D. & Higgs, D. R. *Nucleic Acids Res.* **15**, 6197-6207 (1987).
33. Lalande, M., Noolandi, J., Turmel, C., Rousseau, J. & Slater, G. W. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8011-8015 (1987).
34. Schwartz, D. C. & Cantor, C. R. *Cell* **37**, 67-75 (1984).

ACKNOWLEDGEMENTS. We are grateful to Dr L. Kunkel for his critical reading of the manuscript, D. Shook and K. Glatt for technical assistance, B. Woolf for help in preparation of the manuscript, Dr D. Ledbetter for providing the CMW-1 probe and Dr C. Sapienza for communicating results and hypotheses before their publication. We also thank Dr S. Kapur for clinical evaluation of one of the patients. This work was supported in part by the NIH. M.L. is a senior associate of the Howard Hughes Medical Institute.

Requirement for integrins during *Drosophila* wing development

Danny L. Brower & Sharon M. Jaffe

Department of Molecular and Cellular Biology, Biosciences West Building, University of Arizona, Tucson, Arizona 85721, USA

THE position-specific (PS) integrins of *Drosophila*^{1,2} are highly homologous to vertebrate integrins³⁻⁵, most of which are cell-surface receptors for extracellular matrix components^{6,7}. Integrins are heterodimers, each consisting of noncovalently associated α - and β -subunits. As for the subfamilies of vertebrate integrins, the same β -subunit is found in both *Drosophila* PS integrins, combined with a specific α -subunit to generate either a complete functional PS1 or PS2 integrin^{1,5,8}. Both α - and β -subunits are large transmembrane proteins (relative molecular masses >100,000). Either one or both of these two PS integrins are expressed in most fly tissues during development. A particularly intriguing pattern of expression is found in the mature wing imaginal disc, where the PS1 integrin is expressed primarily on the presumptive dorsal wing epithelium, and the PS2 integrin is found almost exclusively on the ventral epithelium¹. Immediately after pupariation, the central wing pouch evaginates, folding along its centre to appose the epithelia that will secrete the dorsal and ventral surfaces of the adult wing blade⁹. Here we report the results of a genetic analysis indicating that both of the PS integrins are required to maintain the close apposition of the dorsal and ventral wing epithelia during morphogenesis. Also, we conclude that the integrins are not necessary for the maintenance of the cell lineage restriction between the two presumptive wing surfaces in the developing imaginal disc¹⁰⁻¹².

The α -subunit of the PS2 integrin seems to be encoded by the *inflated* (*if*) locus on the X chromosome¹³, and null mutations at *if* cause embryonic lethality (M. Wilcox, A. DiAntonio and M. Leptin, manuscript submitted; see also Fig. 2 legend). We examined *if*³ mutant larval tissues by using monoclonal antibodies directed against the PS integrins (anti-PS antibodies), and our results suggest that the *if*³ allele is a regulatory mutation

at the locus. Homozygous *if*³ late third-instar larvae (just before pupariation and disc evagination) had apparently normal levels of PS2 integrin on most tissues (for example, muscles, salivary glands), but displayed greatly reduced levels of integrin on the surfaces of some imaginal disc cells (Fig. 1*b*). Specifically, there was relatively little PS2 integrin α -subunit in the ventral region of the wing pouch; these cells normally express the protein at very high levels. PS2 integrin expression was not so severely reduced in other regions of the disc, such as the peripodial membrane. The expression of the PS1 integrin in the wing disc seemed to be unaffected by *if*³ (not shown).

Although all of the mutant wing discs examined were clearly affected, there was variability in the extent of PS2 integrin reduction (for example, Fig. 1*b, c*). In the wing pouch, mutant discs typically stained at higher levels anteriorly, and often displayed significant levels of staining with anti-PS2 antibody dorsally (Fig. 1*c*). Similar antero-posterior asymmetry and dorsal staining are characteristic of wild-type PS2 integrin expression in mid-third instar discs¹⁴, about 24 h before pupariation. It is interesting that the pattern of PS2 integrin expression in *if*³ discs seemed relatively normal at the mid-third larval instar (not shown), and it is possible that this allele simply does not allow the cells of the pouch to progress to the more mature pattern of expression.

Adult flies bearing the *if*³ allele display wing blisters, in which the dorsal and ventral surfaces of the wing blade are separated¹³. We found that the penetrance and expressivity of the *if*³ blisters were variable (Fig. 2); this is not surprising in light of the variability of the immunofluorescence results. Nonetheless, these data indicate that the PS2 integrin is important for maintaining the close apposition of the dorsal and ventral epithelia during morphogenesis of the adult wing.

That this is the case was confirmed by clonal analysis studies¹⁵ of flies with mutations in the *mysospheroid* gene (*mys*). The *mys* locus encodes the common PS integrin β -subunit^{4,5}, so null mutations at this locus would be expected to eliminate both PS1 and PS2 integrins⁵. Flies homozygous for null mutations of *mys* die as embryos (see refs 16 and 17 for descriptions of the *mys* lethal phenotype), so to examine the role of integrins in wing morphogenesis, we made clones of cells homozygous for a null mutation of *mys*, along with the flanking markers *yellow* (*y*) and



FIG. 1 Anti-PS2-antibody immunofluorescence of the basal surfaces of late third-instar wing imaginal discs. At this stage, the disc consists primarily of a single layered but highly folded columnar epithelium of ~50,000 cells²². All micrographs are centred on the 'pouch' region, which evaginates into a flat sac before secreting the cuticle of the adult wing blade⁹. Images are reverse-contrast, so staining is dark. *a*, Wild type. PS2 integrin is found at high concentration throughout the ventral (upper) region of the wing pouch. The sharp horizontal boundary of staining across the pouch marks the line along which the epithelium will fold during evagination (that is, the presumptive wing margin). *b*, Mutant for *if*³. PS2 expression is greatly reduced in the pouch. Relatively high levels of staining remain in more peripheral areas, especially along the anterior edge of the pouch (arrow). *c*, Mutant for *if*³. Higher magnification of another disc with a less extreme phenotype, showing residual PS2 in the pouch. Typically, expression is greater in the anterior (left) half of the pouch, and is often seen dorsally (arrow) as well as ventrally. Scale bar, 50 μ m.

METHODS. Mutants for *if*³ were grown as a homozygous stock at 25 °C. Immunofluorescence of wing discs was performed with an anti-PS2 (CF.2C7) monoclonal antibody and fluorescein-conjugated goat anti-mouse antibody (Antibodies Inc.) as previously described¹. Images were detected with an ISIT video camera, and stored on video tape.

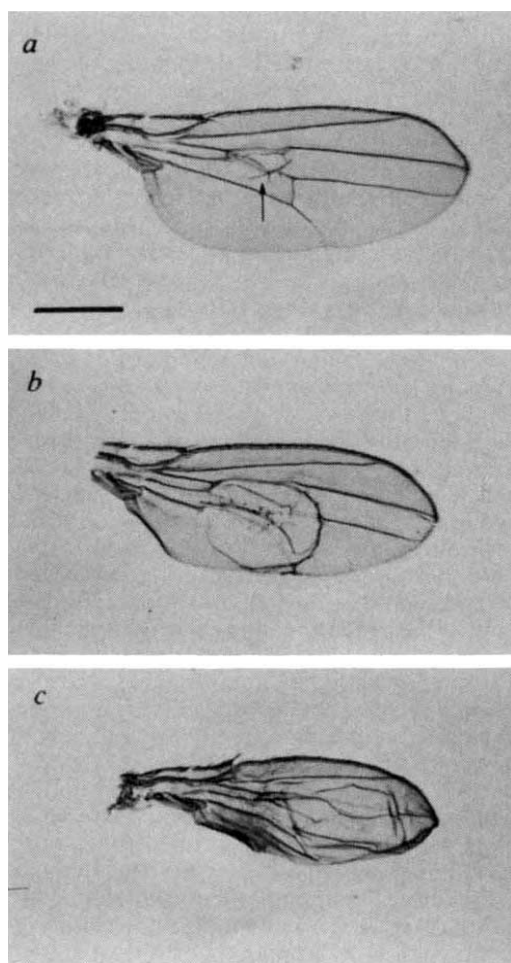


FIG. 2 Adult wings from *if³/if^{k27e}* flies. *a*, A small blister. This phenotype is relatively rare, but when seen, such minor defects are generally found distal to the anterior crossvein. *b*, A large round blister. This is the most common phenotype seen in *if³* homozygotes. *c*, The most extreme phenotype, in which there is a general defect in wing flattening. This phenotype is especially common in *if³/if^{k27e}* flies. Overall, the penetrance of the blister phenotype was variable, but typically at least one wing blister was observed in 15–20% of *if³* homozygotes or hemizygotes (males). The penetrance increased to 60–70% for *if³/if^{k27e}* flies. The penetrance was decreased by low temperature and crowding, and we have noted in most experiments that the first progeny from a cross were much more likely to have blisters than the last flies to eclose. Scale bar, 500 μ m. METHODS. Males (*y w if³*) were crossed to females (*r^{54c} if^{k27e} f/FM7*) to generate *if³/if^{k27e}* flies (identifiable as adults or larvae by their *y⁺* phenotype). Adult wings were stored in ethanol-glycerol (7:3), dehydrated in ethanol, and mounted in Euparal (ASCO Labs) for microscopy. For descriptions of mutations see refs 13, 23 and 24. The alleles *if³* and *if^{k27e}* are inferred to be alleles of the locus encoding the PS2 integrin α -subunit on the basis of the following observations: within the limits of resolution, the genetic map positions of *if³* and *if^{k27e}* correspond to the chromosomal location of the gene for the α -subunit of PS2, the latter based on *in situ* hybridization of molecular probes^{3,13,24}. Flies bearing the *if³* allele express reduced amounts of PS2 integrin in some disc locations (as reported here), and no PS2 integrin is detectable on immunoblots from *if^{k27e}* embryos (M. Wilcox, A. DiAntonio and M. Leptin, manuscript submitted). The *if³* blister phenotype, and the reduction in PS2 expression, are enhanced when *if³* is in *trans* with *if^{k27e}*. Although improbable, it remains a formal possibility that the *if* locus defined by one or both of these alleles encodes a factor that regulates a closely linked gene for the α -subunit of PS2; in any case, this distinction is probably unimportant for the purposes of this study.

forked (*f*), in a phenotypically wild-type background. (see Fig. 3 legend) Because essentially every cell of the wing blade epithelium gives rise to a trichome or bristle which is affected by the marker mutations, the entire structure can be scored for *y mys f* mutant clones with high spatial resolution.

Blisters were the most common phenotype associated with mutant wing clones (Fig. 3*a–c*), and were almost always found wherever there was a large clone (>100 cells). Other phenotypes included folds in one or both surfaces of the wing (more common for long thin clones; Fig. 3*d*), or minor disruptions in the planarity of the wing. The latter were mostly associated with small clones of fewer than 50 cells, and probably represented blisters that simply were too small to allow for a large separation of the epithelia. Blisters were usually much larger than the associated clones, and although the clones generally were elongated in the proximo-distal axis, the blisters tended to be grossly circular in form, like the moderately sized *if* blisters. Ventral and dorsal clones had similar phenotypes (for example, see Fig. 3*a, b*). In general, there was no indication that the *mys* mutation affected the ability of cells to differentiate cuticular structures such as bristles, sensillae, or the thickened cuticle characteristic of veins.

The wing margin was generally unaffected by the absence of PS integrins. Normal morphology was seen around *mys* clones that included many (20 or more) bristles of the anterior triple row (Fig. 3*e*), or hairs of the posterior row (Fig. 3*f*), and the only large wing clones that failed to produce blisters were those confined to the margin region.

The above data demonstrate that the PS integrins are required on both dorsal and ventral epithelia to maintain the close apposition of the wing surfaces. These surfaces first come into contact as a result of the evagination of the wing pouch in the puparium⁹. During the subsequent morphogenesis of the wing, the dorsal and ventral surfaces separate and rejoin at least twice^{9,18}. The first separation is so great that the wing becomes almost circular in cross section; but the two surfaces apparently remain connected by thin basal processes^{9,18}. Considering the role of integrins in focal adhesion sites in vertebrate cells¹⁹, as well as in muscle attachment sites in *Drosophila*^{3,5}, it seems reasonable to hypothesize that the PS integrins are components of the basal attachment sites that connect the pupal wing surfaces¹⁸. The expression of different PS integrins on the two sides of the attachment is reminiscent of the molecular polarity of the muscle attachment sites in the embryo⁵.

It is interesting that the blisters generally were much larger than the associated *mys* clones, and in contrast to the clones, the blisters were roughly circular in shape. This indicates that the physical forces pushing the epithelia apart (probably hydrostatic pressure of the haemolymph) are substantial, and an individual wild-type cell is likely to be torn from its counterpart on the opposite surface if its neighbours are not also joined. We did not detect any consistent spatial relationship between the clones and blisters. That is, clones were found on the inside and outside edges or in the middle of blisters. (The potential interactions of specific mutant effects and the physical constraints inherent in the developing wing are discussed in detail by Waddington⁹.)

The dorso-ventral specificity of PS1 and PS2 integrin expression in the wing disc arises at about the time that a cell lineage restriction is established between these two domains^{14,20}. This correlation indicated that the integrins might be causally related to the lineage restriction. Our data do not support such a relationship. Some *mys* clones did cross the wing margin, although this is to be expected of any clone generated as early as 48 h of development²⁰. More informative are the clones that defined the wing margin for large distances, without crossing to the other surface. For example, the large clone partially illustrated in Fig. 3*f* included 69 dorsal hairs of the marginal posterior row, but no ventral hairs. Similarly, we found clones that included 20 or more bristles of the ventral row or the dorsally

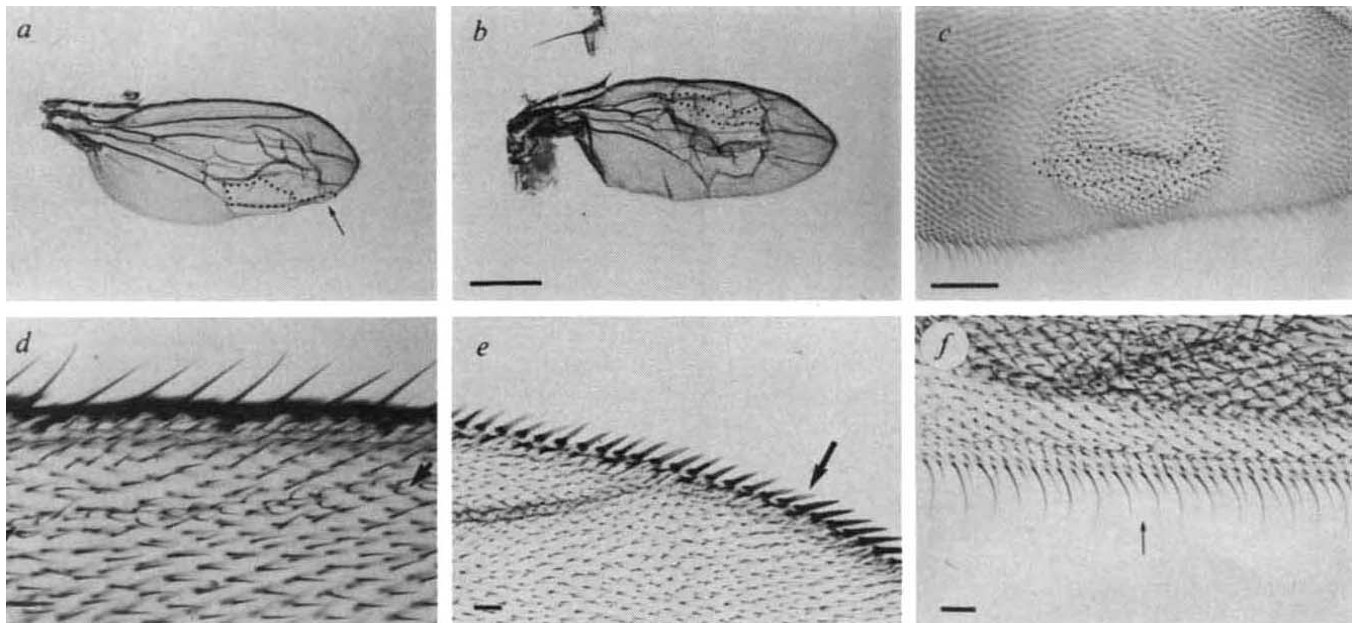


FIG. 3 Homozygous *mys*^{XG43} clones in adult wings. All clones illustrated were generated by irradiation at 48 ± 4 h after egg laying, except for that shown in *c*, which was generated at 72 ± 4 h. *a*, *b*, Blisters caused by large dorsal (*a*) or ventral (*b*) clones. The approximate extent of the clones is indicated by the dotted outlines; the clone in (*a*) abuts the wing margin distally (arrow). Note that the blisters are much larger than the clones, and are roughly circular. *c*, Blister resulting from small clone. The smooth oval shape of the blister contrasts with the irregular shape of the clone (dotted line). *d*, Long thin clone, which results in a discrete fold in the wing (between arrows), in contrast to the blisters. *e*, Clone along the anterior margin. The arrow indicates the last *y f* marginal bristle of the clone; the darker bristles to the right are wild type. This clone includes 35 bristles of the central row, yet has no obvious effect on the morphology of the margin. Overall, the only margin phenotype that was sometimes observed was that of folding near the wing edge, but this seemed to correlate with the presence of nearby blisters, and was probably an artefact of the more general disruption. *f*, Part of a large dorsal clone that abuts the posterior margin. Long ventrally derived wild-type marginal hairs alternate with shorter dorsally derived *y f* hairs (arrow). This clone includes 69 marginal hairs, all dorsal, in addition to its causing a large wing blister. Scale bars: *a* and *b*, 500 μ m; *c*, 100 μ m;

d-f, 10 μ m.

METHODS. For clonal analysis, *y mys*^{XG43 f36a}/FM7 females were crossed to wild-type males. The progeny were grown in uncrowded, well-yeasted vials, and irradiated with X-rays (1,500 rads) at 48 or 72 ± 4 h after egg laying, all at 25 °C. As a control, the same cross was performed with the markers but without the *mys*^{XG43} allele. Embryos homozygous for the *mys*^{XG43} allele²⁵ make no immunologically detectable β -subunit (ref. 5, and R. Salatino, unpublished results), and *mys*^{XG43} behaves as a typical strong allele in various complementation tests (R. Salatino and D.B., unpublished results); we therefore infer that it is a null allele of *mys*. On the basis of sizes and frequencies of *mys*^{XG43} and control clones, there was no indication that *mys*^{XG43} reduced cell proliferation. Initially, 133 wings were scored for marked clones without regard to overall wing morphology. These wings yielded 44 *mys*^{XG43} clones of greater than 50 cells, and virtually all of these clones were associated with clear morphological defects. Subsequently, we examined additional wings selected on the basis of abnormal morphology. Ultimately, a total of 135 wing clones were scored, about half of which included more than 100 cells. Although all results reported here are for the *mys*^{XG43} allele, similar wing blisters have been observed in clones of another putative null allele, *mys*^{XB87} (our unpublished observation).

derived central row of the anterior margin, but which did not cross to the other wing surface. These clones apparently reached the presumptive margin after the lineage restriction was established; even so, clones that define the margin over such distances would not be expected if the integrins are important for the maintenance of the lineage restriction.

Although most third instar imaginal disc cells express some surface PS integrin, these proteins are found in particularly high concentrations on the basal surface of the wing pouch epithelium¹⁴. Also, the spatial specificity of the PS1 and PS2 integrins in this disc region indicates that they are likely to be involved in a specific morphogenetic process there, and we have shown that the integrins are indeed critically important for the proper joining of the dorsal and ventral wing epithelia. It is interesting that our *mys* clones in the leg (15–30 bristles) or abdomen (7–10 bristles) displayed no obvious morphological phenotype (unpublished; see also ref. 21). It seems that the integrins serve a general function in discs, perhaps related to epithelium–matrix connections, but that they additionally perform a much more specialized morphogenetic function in the developing wing. □

Received 12 June; accepted 19 September 1989.

1. Brower, D. L., Wilcox, M., Piovant, M., Smith, R. J. & Reger, L. A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7485–7489 (1984).

2. Leptin, M., Aebersold, R. & Wilcox, M. *EMBO J.* **6**, 1037–1043 (1987).
3. Bogaert, T., Brown, N. & Wilcox, M. *Cell* **51**, 929–940 (1988).
4. MacKrell, A. J., Blumberg, B., Haynes, S. R. & Fessler, J. H. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2633–2637 (1988).
5. Leptin, M., Bogaert, T., Lehmann, R. & Wilcox, M. *Cell* **56**, 401–408 (1989).
6. Hynes, R. O. *Cell* **48**, 549–554 (1987).
7. Ruoslahti, E. & Pierschbacher, M. D. *Science* **238**, 491–497 (1987).
8. Wilcox, M., Brown, N., Piovant, M., Smith, R. J. & White, R. A. H. *EMBO J.* **3**, 2307–2313 (1984).
9. Waddington, C. H. *J. Genet.* **41**, 75–139 (1941).
10. Bryant, P. J. *Dev. Biol.* **22**, 389–411 (1970).
11. Garcia-Bellido, A. & Merriam, J. R. *Dev. Biol.* **24**, 61–87 (1971).
12. Garcia-Bellido, A., Ripoll, P. & Morata, D. *Dev. Biol.* **48**, 132–147 (1976).
13. Lindsley, D. L. & Grell, E. H. *Genetic Variations of Drosophila melanogaster* (Carnegie Institution of Washington Pub. No. 627, 1968).
14. Brower, D. L., Piovant, M. & Reger, L. A. *Dev. Biol.* **108**, 120–130 (1985).
15. Postlethwait, J. H. in *Genetics and Biology of Drosophila*, Vol. 2c (eds Ashburner, M. & Wright, T. R. F.) 359–441 (Academic, London, 1978).
16. Wright, T. R. F. *J. exp. Zool.* **143**, 77–99 (1960).
17. Newman, S. M. & Wright, T. R. F. *Dev. Biol.* **86**, 393–402 (1981).
18. Milner, M. J. & Muir, J. *Wilhelm Roux Arch. dev. Biol.* **196**, 191–201 (1987).
19. Burridge, K., Fath, K., Kelly, T., Nuckolls, G. & Turner, C. A. *Rev. Cell Biol.* **4**, 487–525 (1988).
20. Brower, D. L. *Cell* **41**, 361–364 (1985).
21. Wieschaus, E. & Noell, E. *Wilhelm Roux Arch. dev. Biol.* **195**, 63–73 (1986).
22. Bryant, P. J. in *Genetics and Biology of Drosophila*, Vol. 2c (eds Ashburner, M. & Wright, T. R. F.) 229–335 (Academic, London, 1978).
23. Carlson, P. S. *Genet. Res.* **17**, 53–81 (1971).
24. Falk, D. R., Roselli, L., Curtiss, S., Halladay, D. & Klufas, C. *Mutat. Res.* **126**, 25–34 (1984).
25. Wieschaus, E., Nüsslein-Volhard, C. & Jurgens, G. *Wilhelm Roux Arch. dev. Biol.* **193**, 296–307 (1984).

ACKNOWLEDGEMENTS. We thank Ted Wright, Eric Wieschaus, Darrel Falk, and the stock centres at Indiana and Bowling Green for fly stocks, and Michael Wilcox for communication of unpublished work. We also thank many colleagues in Tucson for criticism of the manuscript. The work was supported by grants from the NSF and NIH.

Requirement of the *Drosophila raf* homologue for *torso* function

Linda Ambrosio*, Anthony P. Mahowald†
& Norbert Perrimon*

* Howard Hughes Medical Institute and Department of Genetics, Harvard Medical School, 25 Shattuck Street, Boston, Massachusetts 02115, USA
† Department of Genetics, Case Western Reserve University, Cleveland, Ohio 44106, USA

IN *Drosophila* the correct formation of the most anterior and posterior regions of the larva, acron and telson is dependent on the maternally expressed terminal class of genes¹. In their absence, the anterior head skeleton is truncated and all the structures posterior to the abdominal segment seven are not formed¹⁻⁶. The protein predicted to be encoded by one of these genes, *torso* (*tor*), seems to be a transmembrane protein with an extracytoplasmic domain acting as a receptor and a cytoplasmic domain containing tyrosine kinase activity⁷. Here we report that another member of the terminal-genes class, *l(1)polehole* (*l(1)ph*)^{2,3,8}, which is also zygotically expressed, is the *Drosophila* homologue of the *v-raf*

oncogene and encodes a potential serine-and-threonine kinase. We also show that functional *l(1)ph* gene product is required for the expression of a gain-of-function *tor* mutant phenotype, indicating that *l(1)ph* acts downstream of *tor*. Together, these results support the idea that the induction of terminal development occurs through a signal transduction system, involving the local activation of the *tor*-encoded tyrosine kinase at the anterior and posterior egg poles, resulting in the phosphorylation of the *l(1)ph* gene product. In turn, downstream target proteins may be phosphorylated, ultimately leading to the regionalized expression of zygotic target genes. Such a process is in agreement with the finding that both *tor* and *l(1)ph* messenger RNAs are evenly distributed.

It has been shown previously that the gene encoding a potential serine-and-threonine kinase with homology to the *v-raf* oncogene⁹⁻¹¹ maps near the *l(1)ph* locus^{12,13}. Molecular analysis of the DNA lesions associated with *l(1)ph* mutations indicates that the *Drosophila raf* homologue, *D-raf*, is in fact the *l(1)ph* gene (Fig. 1). Confirmation of the identity of *D-raf* and *l(1)ph* was obtained by P-element-mediated rescue experiments. A DNA fragment of 4.3 kilobases (kb) containing only the *D-raf* coding sequences rescued both the zygotic and maternal effects of *l(1)ph* mutations (Fig. 1b).

The *D-raf* transcripts are distributed homogeneously

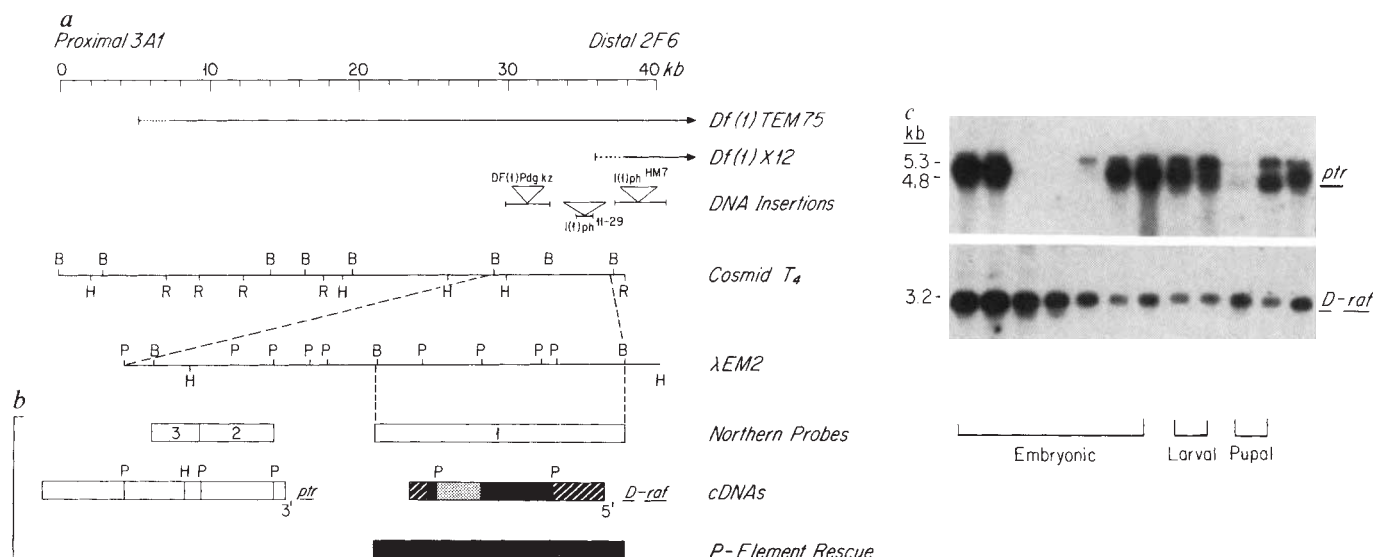


FIG. 1 The molecular organization of the 2F6-3A1 cytogenetic region containing the *l(1)ph* locus. The reference bar is graduated in kb with the position 0 defined by the proximal end of the T4 cosmid²³. a, The *l(1)ph* gene lies between deficiency breakpoints *Df(1)TEM75* and *Df(1)X12*⁸ which map to positions 6–8 and 36–38, respectively. The solid lines represent DNA sequences not removed by the deficiencies and the broken lines represent the maximum limits within which the deficiency breakpoints fall. *Df(1)TEM75* complements *l(1)ph* mutations, defining a proximal limit for the *l(1)ph* gene. The *Df(1)X12* chromosome does not complement *l(1)ph* mutations, indicating at least some of the *l(1)ph* gene lies proximal to the breakpoint at position 36–38. DNA polymorphisms in three different strains are identified as insertions in the *l(1)ph* region. Two mutant alleles of *l(1)ph*, *l(1)ph^{HM7}* and *l(1)ph¹¹⁻²⁹* (L.A., manuscript in preparation) have ~4-kb DNA insertions mapping to positions 38–41 and 35–36, respectively. The *Df(1)Pdgkz* chromosome has wild-type *l(1)ph* activity but contains a 7-kb DNA insertion at position 30–33. These DNA insertions are depicted at their approximate positions by the inverted triangles. The deficiency breakpoints and DNA insertions were mapped using standard methods as described previously²⁴. b, Two non-overlapping transcription units lie within the genomic region containing the *l(1)ph* gene. The *D-raf* gene was localized to position 33–37 using a *Bam*H1 subclone (northern-blot probe 1) of the genomic probe *D-raf1* (ref. 12). A second transcription unit called *ptr* (proximal to *raf*) was identified by northern-blot analysis using probes 2 and 3. No additional bona fide transcription units are encoded within the DNA spanning positions 0–40. Complementary DNAs homologous to *D-raf* and *ptr* transcription units were

isolated from λ gt11 (gift of K. Zinn and C. Goodman) and plasmid pNB50 (ref. 25) cDNA libraries, respectively. The 3.2-kb *D-raf* cDNA contains both 700 base pairs-5' and 600 base pairs-3' nontranslated regions (hatched boxes), and a coding region of 1.9 kb (shaded boxes) which includes the putative serine-threonine kinase domain (vertical striped box). The 4-kb insertion in *l(1)ph¹¹⁻²⁹* maps within the *D-raf* coding region, whereas the *l(1)ph^{HM7}* insertion maps 5' to it. The complete structure of the 4.3 kb *ptr* cDNA has not been determined; but the *Df(1)Pdgkz* DNA insertion occurs within sequences encoding the *ptr* transcript. *Df(1)Pdgkz/Df(1)X12* females show only truncated *ptr* transcripts, have the normal 3.2 kb *D-raf* transcript, and are viable, fertile adults (data not shown). These data indicate that it is the *D-raf* rather than the *ptr* transcription unit which encodes the *l(1)ph* gene. A transformed strain, B-13-1 (ref. 13) containing a P-element construct carrying the wild-type *D-raf* gene was tested for complementation to *l(1)ph* mutations. All pleiotropic aspects of the mutant *l(1)ph* phenotype including the maternal effect, were rescued by the transformed DNA fragment in *trans* configuration (data not shown). c, Accumulation of *ptr* and *D-raf* transcripts during development. The *ptr* transcription unit encodes a 5.3 kb and a 4.3 kb RNA species and its expression is developmentally regulated. The 3.2 kb *D-raf* transcript is present at all stages of development. The time points for the embryonic period are given as hours after egg laying. These northern blots re-probed with an actin 5C probe indicate that each lane contains an equivalent concentration of RNA (data not shown). Preparation of poly(A)⁺ RNA and northern-blot analysis were performed as described previously²⁶.

throughout the oocyte and embryo at all stages (Fig. 2) and are not preferentially localized to the termini. Based on evidence presented below, it is likely that the maternal *D-raf*-encoded protein is also evenly distributed in the embryo. Similarly, there is no specific pattern to the localization of *tor* mRNA in the egg or embryo, and a uniform distribution of *tor* gene product is also predicted⁷. Thus the spatial restriction of both *D-raf* and *tor* activities, and hence the proper determination of terminal cell fates, does not rely on the spatial localization of the corresponding gene products, but could rather depend on their localized activation, possibly by one of the other terminal genes.

The finding that the proteins encoded by *D-raf* and *tor* contain domains with potential kinase activities indicates a possible mechanism by which such restricted activity of evenly distributed gene products occurs. In mammalian cells, the proto-oncogene homologue of *v-raf*, *c-raf-1* (refs 14, 15) encodes a protein which functions as a transducer of extracellular growth signals. Binding of growth factors, such as epidermal growth factor (EGF) and platelet-derived growth factor (PDGF) to receptors with tyrosine kinase activity causes the phosphorylation of *c-raf-1* product and the stimulation of its kinase activity¹⁶. Activated *c-raf-1* protein in turn phosphorylates other proteins, which may include itself and transcription factors, to initiate the programme for cell division. The proteins encoded by *tor* and *D-raf* could function in an analogous signal transduction pathway. The *tor* gene product could be the receptor for a spatially localized 'terminal' signal potentially secreted by the follicle cells during oogenesis. Binding of the localized ligand to the extracellular receptor domain of the *tor* gene product would activate the kinase domain, which would then locally phosphorylate the ubiquitous *D-raf* gene product. Thus, local activation of the *D-raf* gene product would lead to the induction of a programme of zygotic gene expression specific for the termini of the embryo.

We tested genetically for the interactions between *tor* and *D-raf* which such a model predicts. Loss-of-function alleles of *tor* lead to the loss of terminal structures identical to those described for null *l(1)ph* mutations. The maternal effect of some *tor* alleles, however, produce an opposite, gain-of-function embryonic phenotype^{17,18}. Females homozygous for the *tor* gain-of-function allele *tor*^{RL3} produce at 25 °C embryos that lack thoracic and abdominal structures but show differentiation of the terminal Anlagen. Loss of abdominal segments is correlated with the suppression of early gap and segmentation gene expression, indicating that *tor*^{RL3} could encode a protein which is ectopically active in the central domain of *tor*^{RL3}-derived embryos^{17,18}.

If the *D-raf* gene product is required to transfer the activated *tor* signal to other downstream terminal genes, then the absence of *D-raf* should suppress the *tor*^{RL3} gain-of-function phenotype. To test whether this occurs, embryos derived from eggs lacking maternal *D-raf* and containing *tor*^{RL3} activities were examined. Two classes of embryos were observed with equal frequency after 24 hours of development at either 18 °C or 25 °C. Class 1 embryos showed the terminal class defect, with anterior and posterior regions of the embryos failing to develop. These embryos most probably represent the progeny lacking maternal *D-raf* activity. Class 2 embryos suffered massive cell death and developed little cuticle. These embryos are equivalent to those lacking both maternal and zygotic *D-raf* activity. No embryos exhibited the *tor*^{RL3} phenotype. Thus, functional *D-raf* gene product is required for expression of the gain-of-function *tor*^{RL3} phenotype. Loss of *D-raf* gene product blocks the transfer of the *tor* terminal signal in the central domain and at the termini of *tor*^{RL3}-derived embryos. This result also indicates that *D-raf* gene product is present and can be activated in the central domain of *tor*^{RL3} and wild-type embryos, which is consistent with the ubiquitous presence of *D-raf* transcripts (Fig. 2).

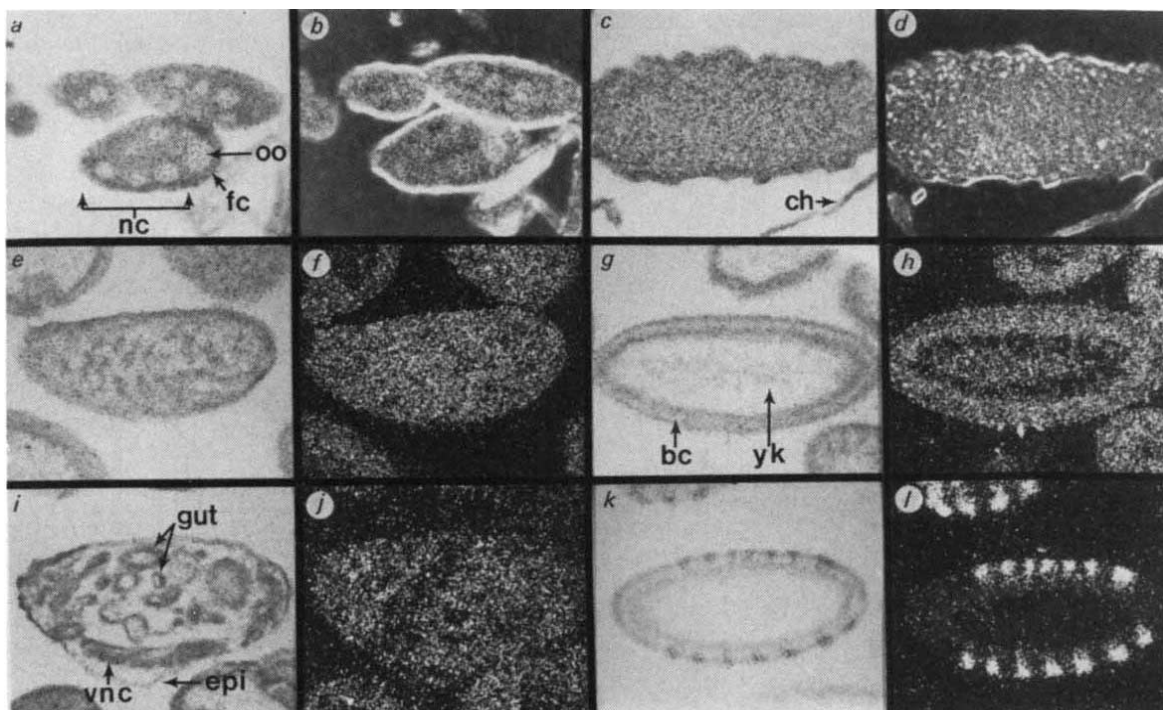


FIG. 2 Accumulation of *D-raf* transcripts during oogenesis and embryogenesis. *a, b*, An equal distribution of *D-raf* transcripts is observed for germ-line-derived nurse cells, oocyte and somatic follicle cells at stages 5–9 of oogenesis. *c, d*, A mature stage-14 oocyte shows no spatial localization of maternal *D-raf* RNAs. *e, f*, Accumulation of *D-raf* RNA in a syncytial blastoderm staged embryo 2 h after fertilization. *g, h*, At 3 h, *D-raf* RNA is present in the peripheral blastoderm cells and in the internal yolk region.

i, j, At later stages of embryogenesis (16 h) *D-raf* transcripts are present in endodermal, mesodermal and ectodermal tissues. *k, l*, Control blastoderm stage embryo showing localization of *fushi tarazu* (*ftz*) transcripts²⁷. Preparation of tissue sections and hybridization conditions using ³⁵S-labelled *D-raf* and *ftz* probes were as described previously²⁶. Abbreviations: bc, blastoderm cells; ch, chorion; epi, epidermis; fc, follicle cells; nc, nurse cell complex; oo, oocyte; vnc, ventral nerve cord; yk, yolk.

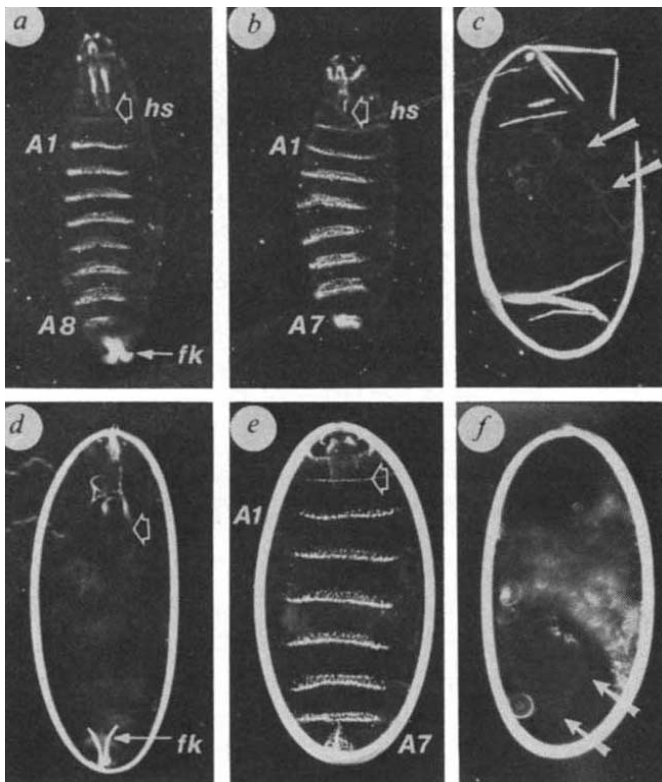


FIG. 3 The phenotype of embryos derived from double mutant *l(1)ph* and *tor*^{RL3} female germ cells. Dark field photographs of cuticular preparations of wild-type embryos (a) and of embryos derived from; homozygous *l(1)ph* female germ cells (b, c); homozygous *tor*^{RL3} females (d); homozygous *l(1)ph;tor*^{RL3} female germ cells (e, f). In b and e, embryos show the terminal class phenotype: a truncated head skeleton and the lack of structures posterior to abdominal segment seven. These embryos developed without maternal *D-raf* activity. In c and f, embryos show the *l(1)ph* 'null' phenotype; little cuticle is formed. These embryos developed without maternal and zygotic *D-raf* activity. d, The *tor*^{RL3} gain-of-function phenotype; at 19 °C embryos form head and tail structures without thoracic or abdominal segments. The *l(1)ph*^{EA75} allele used to generate female germ cells lacking *D-raf* activity as previously described²⁸. All females were crossed to Oregon R wild-type males. Abbreviations: fk, filzkörper; hs, head skeleton.

Suppression of the gain-of-function *tor*^{RL3} phenotype is also observed in embryos derived from eggs containing only half the wild-type amount of *D-raf* gene product. At 18 °C, the proportion of embryos that hatched was greater for these embryos than for those with normal levels of *D-raf* activity (Table 1). The maternal effect of *tor*^{RL3} is cold sensitive¹⁸; at 25 °C, embryos show complete loss of abdominal segmentation, whereas at 18 °C a few embryos (1%) have normal segmentation and hatch. This weak gain-of-function phenotype of the *tor*^{RL3} allele at 18 °C is indicative of a low-level of ectopic *tor* activity. But when the maternal contribution of functional *D-raf* gene product was reduced by half there was a 10-fold increase (11%) in the number of embryos that survive to hatching. The enhanced survival can be attributed to the reduced levels of maternal *D-raf* activity, and, thus, to an indirect reduction of ectopic *tor*^{RL3} activity. When the availability of *D-raf* gene product is reduced, the efficiency of signal transfer may also decrease, thereby restoring normal embryonic development. At the termini, where the activity of *tor*^{RL3} is at least normal, a 50% reduction of *D-raf* product does not adversely affect the development of head or tail structures.

These results support the idea that induction of terminal development acts through a signal transduction system involving a phosphorylation cascade. Spatial restriction of activity is achieved in this pathway by the localized activation of evenly distributed gene products. Two genes likely to participate in this

TABLE 1 Partial suppression of the *tor*^{RL3}

Maternal genotype	T (°C)	N	No. of embryos hatched	Hatching (%)
A. <i>tor</i> ^{RL3} / <i>tor</i> ^{RL3}	18	1,920	20	1
	19	2,812	0	0
	20	1,882	0	0
B. <i>l(1)ph/FM3</i> <i>tor</i> ^{RL3} / <i>tor</i> ^{RL3}	18	1,906	220	11.5
	19	2,788	110	3.9
	20	1,320	3	0.22
C. $\pm$ / <i>F</i> M3: <i>tor</i> ^{RL3} / <i>tor</i> ^{RL3}	18	1,950	30	1.5
	19	1,478	11	0.7
	20	1,520	0	0

The frequency of embryonic hatching was assayed 48 h after egg laying; females were mated to Oregon R wild-type males. A, Eggs produced by *tor*^{RL3}/*tor*^{RL3} females collected from the original stock obtained from T. Schupbach. For B and C the following stock was made: *l(1)ph/FM3/Dp(1;Y)w⁺³⁰³;tor*^{RL3}/*CyO*. From this stock *l(1)ph/FM3;tor*^{RL3}/*tor*^{RL3} females were collected and assayed in B. For the females assayed in C, *l(1)ph/FM3;tor*^{RL3}/*CyO* females were mated to *tor*^{RL3}/*CyO* males and the $\pm$ /*F*M3:*tor*^{RL3}/*tor*^{RL3} females collected. In this experiment, surviving larvae were phenotypically normal and some developed into adults. The unhatched embryos had segmentation defects which were more extreme at 20 °C than at 18 °C. A null allele, *l(1)ph*^{EA75}, was used in this study. Descriptions of balancer chromosomes, *F*M3 and *CyO* in ref. 22.

cascade are *tor* and *l(1)ph* (*D-raf*). We have shown here that *D-raf* acts downstream of the *tor* gene product to transfer the maternal terminal signal. Our data is consistent with the model proposed above and by Sprenger *et al.*⁷: activation of *tor*-encoded tyrosine kinase in a spatially restricted manner results in the phosphorylation of *D-raf* gene product, which in turn phosphorylates serines and threonines of downstream target proteins. One such target could be the protein encoded by the *tailless* (*tl*) gene¹⁹; phosphorylation by *D-raf* gene product of maternally expressed Tll protein¹⁸ could cause its activation, leading directly or indirectly to zygotic expression of *ill*. Other downstream terminal genes, including *fork head*²⁰ and *spalt*²¹ could play a part in mediating the terminal class signal transmitted by *D-raf*.

In contrast to the other maternal terminal class genes, *l(1)ph* gene activity is also required at later embryonic and larval stages of development for the maintenance of cell viability³ and for the proliferation of somatic cells², respectively. Thus the *l(1)ph* gene product acts in several different developmental pathways; only its maternal role in the terminal class pathway has been addressed here. □

Received 10 July; accepted 29 September 1989.

- Nüsslein-Volhard, C., Frohnhofer, H. G. & Lehmann, R. *Science* **238**, 1675–1681 (1987).
- Perrimon, N., Engstrom, L. & Mahowald, A. P. *Dev. Biol.* **110**, 480–491 (1985).
- Ambrosio, L., Mahowald, A. P. & Perrimon, N. *Development* **106**, 145–158 (1989).
- Degelman, A., Hardy, P. A., Perrimon, N. & Mahowald, A. P. *Dev. Biol.* **115**, 479–489 (1986).
- Perrimon, N., Mohler, D., Engstrom, L. & Mahowald, A. P. *Genetics* **113**, 695–712 (1986).
- Schupbach, T. & Wieschaus, E. *Wilhelm Roux Arch. dev. Biol.* **195**, 302–317 (1986).
- Sprenger, F., Stevens, L. M. & Nüsslein-Volhard, C. *Nature* **338**, 478–483 (1989).
- Perrimon, N., Engstrom, L. & Mahowald, A. P. *Genetics* **108**, 559–572 (1984).
- Rapp, U. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 4218–4222 (1983).
- Jansen, H. *et al. Nature* **307**, 281–284 (1984).
- Moelling, K., Heimann, B., Beilmann, P., Rapp, U. R. & Saunders, T. *Nature* **312**, 558–561 (1984).
- Mark, G. E., Macintyre, R. J., Digan, M. E., Ambrosio, L. & Perrimon, N. *Molec. cell. Biol.* **7**, 2134–2140 (1987).
- Nishida, Y. *et al. EMBO J.* **7**, 775–781 (1988).
- Bonner, T. I. *et al. Nucleic Acids Res.* **14**, 1009–1015 (1986).
- Rapp, U. R., Cleveland, J. L., Storm, S. M., Beck, T. W. & Huleihel, M. in *Oncogenes and Cancer* (ed. Aaronson, S. A.) 55–74 (Japan Sci. Soc., Tokyo, 1987).
- Morrison, D. K., Kaplan, D. R., Rapp, U. R. & Roberts, T. M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8855–8859 (1988).
- Klinger, M., Erdelyi, M., Szabad, J. & Nüsslein-Volhard, C. *Nature* **335**, 275–277 (1988).
- Strecker, T. R., Halsell, S. R., Fisher, W. W. & Lipshitz, H. D. *Science* **243**, 1062–1066 (1988).
- Strecker, T. R., Merriam, J. R. & Lengyel, J. A. *Development* **102**, 721–734 (1988).
- Weigel, D., Jürgens, G., Kuttner, F., Seifert, E. & Jäckle, H. *Cell* **57**, 645–658 (1989).
- Frei, E. *et al. EMBO J.* **7**, 197–204 (1988).
- Lindsley, D. L. & Grell, E. H. in *Genetic variations of Drosophila melanogaster* (Carnegie Institution of Washington, Publication No. 627, 1968).

23. Mariani, C., Pirrotta, V. & Manet, E. *EMBO J.* **4**, 2045–2052 (1985).
 24. Maniatis, T., Fritsch, E. F. & Sambrook, J. in *Molecular Cloning: a Laboratory Manual* (Cold Spring Harbor Press, New York, 1982).
 25. Brown, N. H. & Kafatos, F. C. *J. molec. Biol.* **203**, 425–437 (1988).
 26. Shi-Chung N. G., Perkins, L. A., Conboy, G., Perrimon, N. & Fishman, M. C., *Development* **105**, 629–638 (1989).
 27. Hafen, E., Kuroiwa, A. & Gehring, W. J. *Cell* **37**, 833–841 (1984).
 28. Perrimon, N. *Genetics* **108**, 927–939 (1984).

ACKNOWLEDGEMENTS. We thank V. Pirrotta for T4 cosmid DNA, W. Gehring for *ftz* DNA, K. Zinn and N. Brown for cDNA libraries, Y. Nishida for the B-13-1 stock, T. Schubach for the *tor^{RL3}* stock, L. Perkins for the northern blot, and D. Smouse for discussion and comments on the manuscript. This work was supported by grants from the Howard Hughes Medical Institute and the NIH.

Substantial increase of protein stability by multiple disulphide bonds

Masazumi Matsumura*, Giovanni Signor & Brian W. Matthews

Institute of Molecular Biology and Department of Physics, University of Oregon, Eugene, Oregon 97403, USA

DISULPHIDE bonds can significantly stabilize the native structures of proteins^{1–3}. The effect is presumed to be due mainly to a decrease in the configurational chain entropy of the unfolded polypeptide^{4–7}. In phage T4 lysozyme, a disulphide-free enzyme, engineered disulphide mutants that crosslink residues 3–97, 9–164 and 21–142 are significantly more stable than the wild-type protein^{8–11}. To investigate the effect of multiple-disulphide bonds on protein stability, mutants were constructed in which two or three stabilizing disulphide bridges were combined in the same protein. Reversible thermal denaturation shows that the increase in melting temperature resulting from the individual disulphide bonds is approximately additive. The triple-disulphide variant unfolds at a temperature 23.4 °C higher than wild-type lysozyme. The results demonstrate that a combination of disulphide bonds, each of which contributes to stability, can achieve substantial overall improvement in the stability of a protein.

Unpaired cysteine residue(s) in a protein containing a disulphide bond(s) can lead to oligomerization through thiol/disulphide interchange¹². Previously described single-disulphide mutants (designated as 3C–54T, 9C–164C-wt* and 21C–142C-wt*, see Fig. 1) were, therefore, constructed in an otherwise cysteine-free pseudo wild-type lysozyme (wt*)^{11,13}. For the same reason the two double-disulphide mutants D3–97/9–164 and D9–164/21–142 and the triple disulphide mutant T3–97/9–164/21–142 were also designed and constructed to have no unpaired cysteines. Figure 1 shows the locations of the disulphide bonds and ancillary mutations introduced into T4 lysozyme. Each of the mutant genes was expressed in *Escherichia coli* and the proteins purified to homogeneity. Immediately after purification, the mutant enzymes were found to be a mixture of oxidized (crosslinked) and reduced (noncrosslinked) forms, as judged by ion-exchange high-performance liquid chromatography (HPLC)¹¹, reversed-phase HPLC^{8,13} and titration of protein thiols with Ellman's reagent (data not shown). After exposure of proteins to air under mild alkaline conditions (pH 8.0) for several days, however, the mutant proteins were all converted to the oxidized forms (Table 1). Nonreducing SDS-PAGE showed that the double- and triple-disulphide mutants migrate faster than the disulphide-free wild-type lysozyme, whereas all the multiple-disulphide mutants had mobility identical with wild type in the presence of reducing agent (data not shown). This

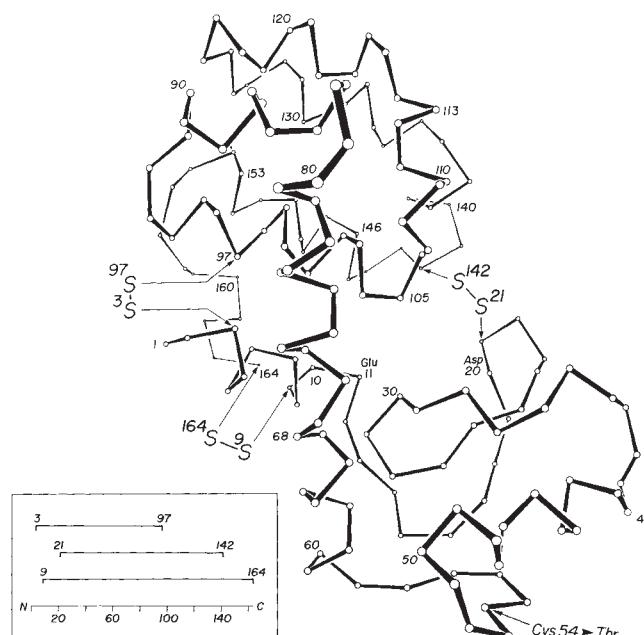


FIG. 1 Backbone of T4 lysozyme showing the locations of the three engineered disulphide bridges. The insert illustrates the loops formed by these bridges. Identification and generation of the mutant lysozymes is as follows:

Variant	Amino-acid replacements						No. of cysteines	No. of disulphide bonds
	3	9	21	54	97	142		
wt	Ile	Ile	Thr	Cys	Cys	Thr	2	0
wt*				Thr	Ala		0	0
3C–54T	Cys			Thr			2	1
9C–164C-wt*		Cys		Thr	Ala	Cys	2	1
21C–142C-wt*			Cys	Thr	Ala	Cys	2	1
D3–97/9–164	Cys	Cys		Thr		Cys	4	2
D9–164/21–142		Cys	Cys	Thr	Ala	Cys	4	2
T3–97/9–164/21–142	Cys	Cys	Cys	Thr		Cys	6	3

Recombinant DNA techniques used in the construction of the disulphide mutants were essentially as described¹⁷. Mutagenic oligonucleotides (21–23 mer) were synthesized using a model 380B DNA synthesizer (Applied Biosystems) and purified by a C18 Sep-Pak cartridge (Millipore). The single-stranded DNA template for site-directed mutagenesis was an M13mp18 derivative containing the T4 lysozyme gene on a 630 base-pair *Bam*HI–*Hind*III fragment. The mutagenesis was performed according to Kunkel *et al.*¹⁸ using *E. coli* strain CJ236 (*dut1, ung1, thi1, relA1/pCJ105* (Cm^r)). After the repair synthesis of DNA *E. coli* JM101 (ref. 19) was transformed and the mutants identified either by DNA sequencing²⁰ or plaque hybridization with the [³²P]-labelled mutagenic prime²¹. The mutations were verified by sequencing the entire T4 lysozyme gene. The mutated T4 lysozyme gene on M13 was digested with *Bam*HI and *Hind*III, and cloned into the expression plasmid pHSe5 that contains *tac* and *lacUV5* promoters, *lacI^q* gene as well as *trp* terminator²². *E. coli* RR1 (ref. 17) was then transformed by the recombinant pHSe5 and the mutant protein was overproduced by addition of isopropyl- β -thiogalactoside. The mutant proteins were purified to homogeneity by CM-Sepharose and SP-Sephadex (Pharmacia) chromatography as described²².

observation indicates that the crosslinked proteins have a more compact structure in the denatured state.

The activity of the D3–97/9–164 mutant (Table 1) is indistinguishable from that of wild-type enzyme in both the oxidized and reduced forms. The result is consistent with the observation that the corresponding single-disulphide mutants have essentially the same activities as wt* in both oxidized and reduced forms¹¹. In addition it indicates that virtually all the D3–97/9–164 molecules have the correct pairing of the disulphide bridges as any mispairing would presumably lead to a loss of activity. The wt* lysozyme loses enzymatic activity at ~55 °C at pH 7.4, whereas the D3–97/9–164 mutant, which is more thermostable than wt* by 15.7 °C at pH 2.0 (see below), retains activity up to

* Present address: Department of Immunology, Research Institute of Scripps Clinic, 10666 North Torrey Pines Road, La Jolla, California 92037, USA.

TABLE 1 Properties of mutant lysozymes

Variant	Ox/Red	Thiols (mol SH/mol prot.)	Relative activity (%)	T_m (°C)	$\Delta T_m(\text{mut-wt}^*)$ (°C)	$\Delta T_m(\text{ox-red})$ (°C)
wt	Ox	1.9	(100)		—	—
	Red	2.0	100	41.9		
wt*	Ox	0.0	(103)			
	Red	0.0	103	41.9	0.0	0.0
3C/54T	Ox	ND	96	46.7	4.8	6.7
	Red	ND	90	40.0	-1.9	
9C/164C/wt*	Ox	ND	106	48.3	6.4	12.9
	Red	ND	99	35.4	-6.5	
21C/142C/wt*	Ox	ND	0	52.9	11.0	13.7
	Red	ND	68	39.2	-2.7	
D3-97/9-164	Ox	0.0	95	57.6	15.7	26.0
	Red	4.1	100	31.6	-10.3	
D9-164/21-142	Ox	0.0	0	58.9	17.0	22.9
	Red	4.1	58	36.0	-5.9	
T3-97/9-164/21-142	Ox	0.2	0	65.5	23.4	31.9
	Red	6.0	43	33.4	-8.5	

Oxidized (crosslinked) mutant lysozymes were prepared on exposure to air *in vitro* by incubating in 0.1 M Tris-HCl, 0.15 M NaCl (pH 8.0) for several days. The reduced (noncrosslinked) forms of the mutant lysozymes were prepared by treating the purified proteins ($\sim 1 \text{ mg ml}^{-1}$) with 6 M guanidine hydrochloride, 20 mM DTT and 1 mM EDTA in 50 mM Tris-HCl (pH 8.3). After incubation for 4 h at 23 °C, the mixture was extensively dialysed under a nitrogen gas purge at 4 °C against 0.2 M NaCl and 1 mM EDTA (pH 2.0). Under the nitrogen atmosphere and at acid pH, the reduced proteins were stable for at least a week at 4 °C. Thiol content was determined by Ellman titration²³. A 0.5 ml fraction of the oxidized or reduced protein (0.02 μmole) was mixed with 2.5 ml of a solution containing 2% SDS, 0.08 M sodium phosphate (pH 8.0), and 0.5 mg ml^{-1} EDTA. To 3 ml of the solution was immediately added 0.1 ml 5,5-dithiobis (2-nitrobenzoic acid) (DTNB) (40 mg DTNB in 10 ml of 0.1 M sodium phosphate, pH 8.0). After 15 min, the absorbance at 410 nm was measured. Lysozyme activities were measured at 23 °C using the turbidity assay²⁴. Protein concentrations were determined spectrophotometrically (molar absorption coefficient for wild-type T4 lysozyme $\epsilon_{280\text{nm}}^{0.1\%} = 1.28$) (ref. 24). T_m is the temperature at the midpoint of the thermal denaturation transition at pH 2.0. $\Delta T_m(\text{mut-wt}^*)$ and $\Delta T_m(\text{ox-red})$ are the differences between mutant and wt* lysozyme and between the oxidized and reduced mutant protein, respectively. Data for wild type and single S-S bridge lysozymes are from refs 10, 11 and 25. Thermal stability was assessed at pH 2.0 by measuring the circular dichroism at 223 nm as a function of temperature^{26,27}. Solutions contained $\sim 0.02 \text{ mg ml}^{-1}$ protein, 0.15 M KCl and 1 mM EDTA, adjusted to pH 2.0 with HCl. To avoid air oxidation, the experiments were performed in a nitrogen atmosphere and the KCl-EDTA solutions were extensively bubbled with nitrogen gas before use²⁸.

$\sim 65^\circ\text{C}$ (data not shown). Mutants D9-164/21-142 and T3-97/9-164/21-142, which include the 21-142 disulphide bond, have virtually no catalytic activity in the oxidized form. This is expected as the 21-142 disulphide bond spans the 'mouth' of the active-site cleft (Fig. 1) leading to the complete abolition of lysozyme activity¹³.

The melting temperatures of the single and multiple disulphide bridge mutants at pH 2.0 are summarized in Table 1 and Fig. 2. Both the double mutants are more heat-stable than their constituent single mutants. In the case of the triple mutant the increase in melting temperature (ΔT_m) is substantial (23.4 °C) and agrees surprisingly well with the sum of the ΔT_m s of the corresponding single disulphide variants (22.2 °C). Measurement of the melting temperature at higher pH values is difficult because the solutions approach boiling point before the unfolding transition is complete. As a guide, however, the estimated reversible melting temperature of the triple disulphide mutant at pH 5.0 is 86 °C, which is 20 °C higher than wild type at the same pH.

It is presumed that a consequence of a disulphide bridge is to reduce the configurational backbone chain entropy of the protein in the denatured state, thus increasing the stability of the folded protein³⁻⁷. Theoretical considerations suggest that the entropic contribution to the free energy of stabilization increases in proportion to the logarithm of the number of residues in the loop formed by the disulphide bridge³⁻⁶. When multiple, overlapping disulphide bonds are introduced into a protein, their effects on the configuration of the unfolded protein are interdependent, and the determination of the entropic term becomes more complicated^{6,7,14}. Nevertheless, reasonable agreement between experiment and theory has been observed for ribonuclease A^{6,14}, ribonuclease T1 (ref. 3) and hen egg-white lysozyme¹⁵. Based on theory,¹⁴ the reductions in entropy in the unfolded protein caused by the single 3-97, 9-164 and 21-142 bridges are, respectively, 20.5, 21.9 and 21.2 entropy units. The sum of these three terms, considered independently, is 63.6 entropy units, which corresponds to 20.0 kcal mol^{-1} at 42 °C, the melting temperature of wild-type lysozyme at pH 2.0. By

contrast, the theoretically estimated¹⁴ entropy reduction when all three bridges are included in the same protein is 56.7 entropy units, corresponding to 17.9 kcal mol^{-1} . Thus, even though the loops formed by the three disulphide bridges have substantial overlap (Fig. 1), this has only a relatively small effect on the

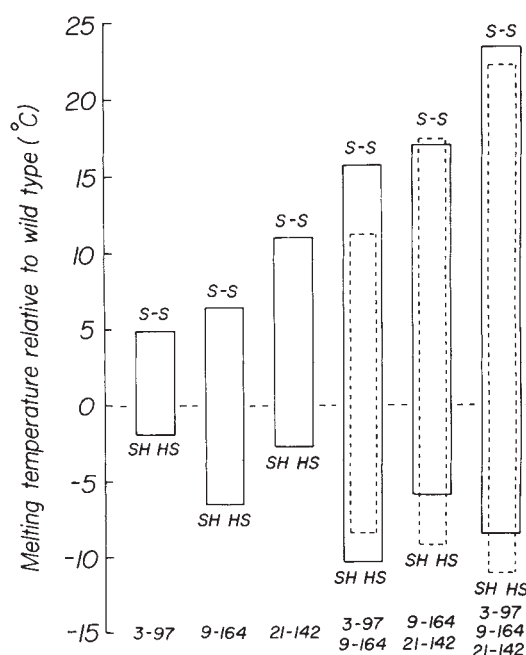


FIG. 2 Melting temperatures of single, double and triple-disulphide bonded lysozymes relative to wild-type lysozyme (ΔT_m) at pH 2.0. The solid bars show the observed melting temperatures of the oxidized and reduced forms of the mutant lysozymes (Table 1). The broken bars for the multiply bridged proteins correspond to the sums of the ΔT_m s for the constituent singly bridged lysozymes.

overall entropy term. Although each disulphide bridge reduces the conformational freedom of the unfolded chain, this restriction represents only a small fraction of the total number of conformational states that are available. Thus, the contributions to stability from each of the three disulphide bridges are still largely additive. The crosslinked lysozymes display a correlation between the number of disulphide bonds and the Stokes radii of the denatured proteins, as judged by gel filtration HPLC in the presence of urea¹⁶. This also indicates that the crosslinks successively decrease the configurational freedom of the unfolded protein.

The above arguments are based on entropy considerations alone. As discussed in ref. 11, the net stabilization of a given disulphide bond depends on other factors, such as the loss of stabilizing interactions associated with introduction of the cysteine(s) and possible introduction of strain associated with the closure of the disulphide bridge. In the case of T4 lysozyme, factors that seem to be helpful in the design of the stabilizing

disulphide bridges include the use of large loops to maximize the entropic effect on the unfolded state and the choice of flexible sites to avoid the introduction of strain into the folded protein¹¹. Exploration of the energetic and theoretical basis of the observed stabilization of the multiply bridged lysozymes will require detailed thermodynamic analysis of these proteins.

In terms of the goal of developing general methods to increase the stability of proteins, the present results are especially encouraging. They demonstrate that individual S-S bridges can be combined together to dramatically enhance stability, even in cases where the disulphide bridges form overlapping loops. It is also encouraging to find that the double-disulphide mutant D3-97/9-164 is fully active and also retains activity at temperatures substantially higher than wild-type lysozyme. This shows that a protein does not have to be marginally stable in order to be fully active. Instead, the present results indicate that the stability of a protein can be substantially increased with no loss in enzymatic activity. □

Received 5 July; accepted 27 September 1989.

- Creighton, T. E. *BioEssays* **8**, 57-63 (1988).
- Thornton, J. M. *J. molec. Biol.* **151**, 261-287 (1981).
- Pace, C. N., Grimsley, G. R., Thomson, J. A. & Barnett, B. J. *J. biol. Chem.* **263**, 11820-11825 (1988).
- Schellman, J. A. *C. R. Trav. Lab. Carlsberg Ser. Chim.* **29**, 230-259 (1955).
- Flory, P. J. *J. Am. Chem. Soc.* **78**, 5222-5235 (1956).
- Poland, D. C. & Scheraga, H. A. *Biopolymers* **3**, 379-399 (1965).
- Chan, H. S. & Dill, K. A. *J. chem. Phys.* **90**, 492-509 (1988).
- Perry, L. J. & Wetzel, R. *Science* **226**, 555-557 (1984).
- Wetzel, R., Perry, L. J., Baase, W. A. & Becktel, W. J. *Proc. natn. Acad. Sci. U.S.A.* **85**, 401-405 (1988).
- Matsumura, M., Becktel, W. J. & Matthews, B. W. *Nature* **334**, 406-410 (1988).
- Matsumura, M., Becktel, W. J., Levitt, M. & Matthews, B. W. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6562-6566 (1989).
- Perry, L. J. & Wetzel, R. *Biochemistry* **25**, 733-739 (1986).
- Matsumura, M. & Matthews, B. W. *Science* **243**, 792-794 (1989).
- Lin, S. H., Konishi, Y., Denton, M. E. & Scheraga, H. A. *Biochemistry* **23**, 5504-5512 (1984).
- Johnson, R. E., Adams, P. & Rupley, J. A. *Biochemistry* **17**, 1479-1484 (1978).
- Signor, G., Matsumura, M., Schellman, J. A. & Matthews, B. W. in *Current Research in Protein Chemistry* (ed. Villafranca, J. J.) (Academic, in the press).

- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
- Kunkel, T. A., Roberts, J. D. & Zakour, R. A. *Meth. Enzym.* **154**, 367-382 (1987).
- Messing, J. *Meth. Enzym.* **101**, 20-78 (1983).
- Sanger, F., Nickelsen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
- Zoller, M. J. & Smith, M. *Meth. Enzym.* **100**, 468-500 (1983).
- Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E. & Dahlquist, F. W. *Meth. Enzym.* **177**, 44-73 (1989).
- Habeeb, A. F. S. A. *Meth. Enzym.* **25**, 457-464 (1972).
- Tsugita, A., Inouye, M., Terzaghi, E. & Streisinger, G. *J. biol. Chem.* **243**, 391-397 (1968).
- Matthews, B. W., Nicholson, H. & Becktel, W. J. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6663-6667 (1987).
- Elwell, M. & Schellman, J. *biochem. biophys. Acta* **386**, 309-323 (1975).
- Hawkes, R., Grütter, M. G. & Schellman, J. *J. molec. Biol.* **175**, 195-212 (1984).
- Becktel, W. J. & Baase, W. A. *Biopolymers* **26**, 619-623 (1987).

ACKNOWLEDGEMENTS. We thank J. Wozniak for technical assistance, F. W. Dahlquist for the lysozyme expression system, T. A. Kunkel for providing *E. coli* strain CJ236 and J. A. Schellman and M. Lindorfer for facilities and advice. Supported in part by grants from the NIH to B.W.M. (and J. A. Schellman), and the Lucille P. Markey Charitable Trust.

Focal points for chromosome condensation and decondensation revealed by three-dimensional *in vivo* time-lapse microscopy

Yasushi Hiraoka, Jonathan S. Minden, Jason R. Swedlow*, John W. Sedat & David A. Agard

Department of Biochemistry and Biophysics, and The Howard Hughes Medical Institute, and *Graduate Group in Biophysics, University of California, San Francisco, San Francisco, California 94143-0448, USA

ALTHOUGH the dynamic behaviour of chromosomes has been extensively studied in their condensed state during mitosis, chromosome behaviour during the transition to and from interphase has not been well documented. Previous electron microscopic studies suggest that chromosomes condense in a non-uniform fashion at the nuclear periphery^{1,2}. But chromosome condensation is a complicated and dynamic process and requires continuous observation in living tissues to be fully understood. Using a recently developed three-dimensional time-lapse fluorescence microscopy technique³, we have observed chromosomes as they relax from telophase, through interphase, until their condensation at the next prophase. This technique has been improved to produce higher-resolution images by implementing new stereographic projection and computational processing protocols⁴. These studies have revealed that chromosomal regions on the nuclear envelope, distinct from the centromeres and telomeres, serve as foci for the decondensation and condensation of diploid chromosomes. The relative positions of the late decondensation sites at the beginning of interphase

appear to correspond to the early condensation sites at the subsequent prophase.

In order to study the chromosomal events that occur between telophase and the ensuing prophase, we have followed embryonic nuclei of *Drosophila melanogaster* during the late syncytial blastoderm period. Chromosomes in living *Drosophila* embryos were visualized by imaging chromatin which had assembled fluorescent-labelled histones into nucleosomes. Calf thymus histones H2A and H2B were conjugated with rhodamine and microinjected into *Drosophila* embryos 30 minutes after oviposition to ensure a uniform distribution of the labelled histones by the time of observation. The injected rhodamine-labelled histones are stably incorporated into chromatin during periods of DNA replication. Under low levels of illumination, these embryos can be observed for more than three hours and develop normally to hatching.

Observing the behaviour of an entire nucleus requires a three-dimensional analysis over time. Imaging at low light levels was accomplished using a cooled, scientific grade charge-coupled device (CCD) imager^{5,6} operated by a self-contained computer workstation. Three-dimensional data sets were obtained by digitally recording a series of five optical sections taken at 1 µm focal intervals every five seconds, as described in the legend to Fig. 1. By repeating this optical sectioning protocol every 25 seconds, a time-lapse series of 80 three-dimensional data sets was recorded. The period of analysis spanned nuclear cycles 12 and 13 of syncytial blastoderm stage embryos. During this stage of embryonic development, the nuclei form an almost synchronously replicating monolayer on the embryo surface^{7,8}. The nuclear cycle length for the embryo shown in Fig. 1 was 22 minutes at 23 °C. Embryos were monitored after data collection to verify that gastrulation occurred normally.

To facilitate analysis of the large amount of three-dimensional data over time, a time-lapse stereo movie was constructed using

the synthetic projection method⁴ described in the legend to Fig. 1. Briefly, a stereo pair of projected images is formed from each set of optical section data by sequentially shifting each image either left or right and summing to generate the projections. The projected images are then computationally processed to remove out-of-focus information⁴. For projected images, this is a simple and straightforward process which can provide nearly exact deblurring. Furthermore, as no light is rejected in the image processing, unlike confocal microscopy, this approach is ideal for the low-light level imaging required for *in vivo* analysis. The resultant in-focus stereo projection images were sequentially displayed in rapid succession as a time-lapse movie. Examples of selected frames of the movie are shown as stereo pairs in Fig. 1.

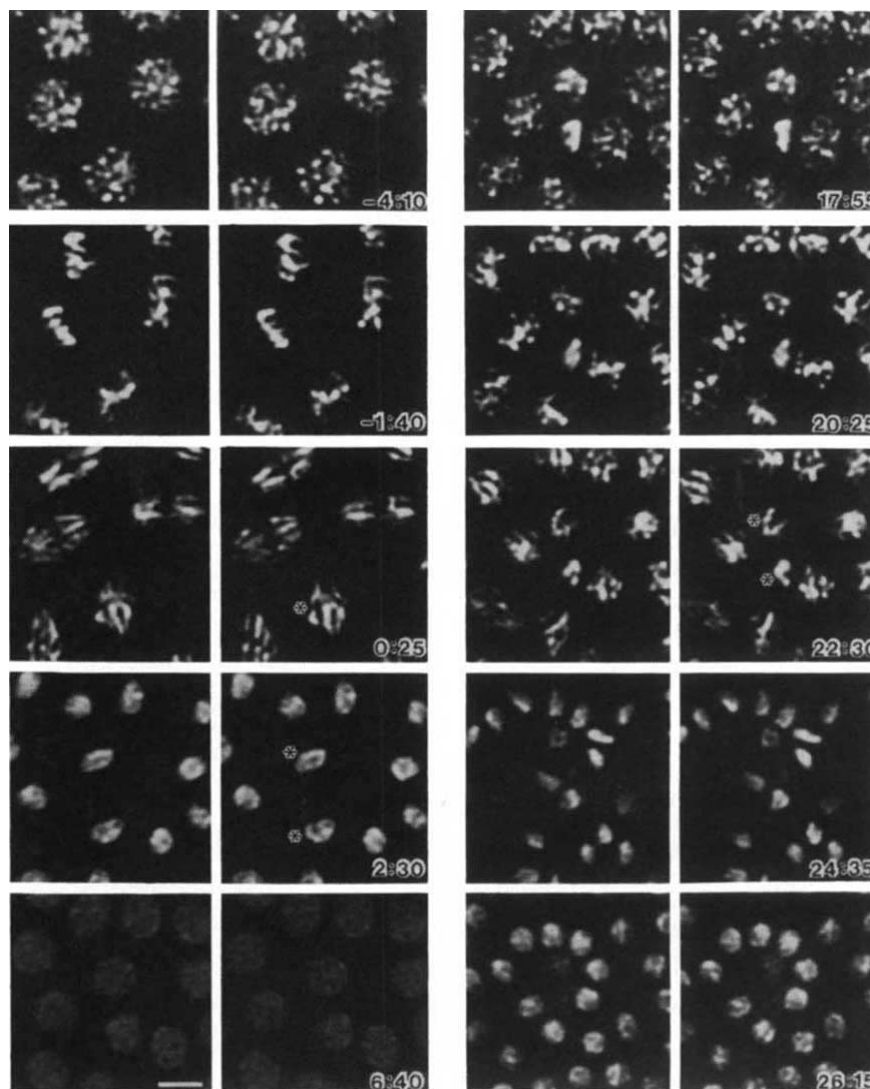
Examination of the real-time, three-dimensional data revealed that chromosome decondensation terminates and condensation initiates at discrete sites predominantly on the nuclear envelope (visualized as bright spots; Fig. 2a-e). Chromosomes begin to decondense shortly after chromatid separation, and by the fol-

lowing interphase, nuclei labelled with rhodamine-conjugated histones show an essentially uniform appearance. Surprisingly, examination of the kinetics of chromosome decondensation revealed that a small portion of the chromosomes remained condensed at several sites on the nuclear envelope (2-3 sites per nucleus) after most of the chromosomes had decondensed (about 5 minutes after the metaphase-anaphase transition of the twelfth cycle). The location of these sites can be seen clearly at the nuclear periphery (indicated by arrows in Fig. 2a-e, column 1). Because these late decondensation sites eventually disappear to make a uniformly stained nucleus during interphase (about 6 minutes after the metaphase-anaphase transition; see Fig. 1, frame at 6:40 min), they are not consistent with the classical definition of heterochromatin.

At the transition from interphase to prophase, bright spots of chromosome condensation were also observed at discrete sites on the nuclear envelope (indicated by arrows in Fig. 2a-e, columns 3-6). The first of these condensation sites appears 10 minutes after nuclear division, with new sites continuing to

FIG. 1 Stereo pairs selected from the time-lapse data. The numbers represent time in minutes:seconds after the 12th nuclear division. Time is stated relative to the onset of chromosome separation at the metaphase-anaphase transition; a negative sign is given to data sets taken before the 12th nuclear division. Asterisks indicate damaged nuclei (see text and also Fig. 2). Scale bar represents 5 μ m.

METHODS. Recording three-dimensional data from living embryos: A standalone, computer workstation CCD system was used which provides real-time capability with a lower digital resolution (12 bits versus 14 bits) but a faster readout rate (200,000 pixel per second versus 50,000 pixel per second) than our previously reported system⁵. All aspects of microscopy, data collection, data storage, image processing and display can be performed on the one workstation. A Peltier-cooled CCD camera (Photometrics, Tucson, Arizona) with a 1340 $\times$ 1037 pixel CCD chip (Kodak-Videk) coated to improve short-wavelength response⁶, is attached to an IMT-2 Olympus inverted microscope. Microscope shutter, focus movement and CCD data collection are controlled by a MicroVax III workstation coupled to a 20 MFLOPS Mercury Zip 3232+ array processor (Mercury Computer Systems, Lowell, Massachusetts) and a Parallax model 1280 graphic display system having 12 Mbyte of image memory (Parallax Graphics, Santa Clara, California). Calf thymus histones were bound to double-strand DNA cellulose to form nucleosomes, labelled with 5(6)-carboxytetramethyl rhodamine *N*-hydroxysuccinimidyl ester (Molecular Probes, Eugene, Oregon), and eluted with a salt gradient. Rhodamine-labelled histones H2A and H2B, at 0.4 mg ml⁻¹, were microinjected into *D. melanogaster* embryos. Details of the method are reported elsewhere³. The rhodamine-histone injected embryos were illuminated with a 100W mercury arc lamp attenuated 20-fold by a neutral density filter and observed using an Olympus 40 $\times$ /NA1.3 oil immersion objective lens and a high selectivity rhodamine filter combination (Omega Optical, Brattleboro, Vermont). To speed up the image readout, images were taken on a small portion of the CCD, 256 $\times$ 256 pixels corresponding to 44 $\times$ 44 μ m area with a pixel size of 0.17 μ m. Five optical section images were recorded at an interval of 1 μ m with an exposure time of 0.1 s. It took 5 s to read out correct and store each image; thus, 25 s for one cycle of a focus series. This was repeated 80 times to form a time course. Generation of a time-lapse stereo movie: To make a pair of stereo projections, each set of 5 optical sections was projected to form a single image for each eye. While forming the projection for the left-eye, images were sequentially shifted by 1 pixel to the left, whereas right eye images were shifted 1 pixel to the right. As the optical sections are separated by 1 μ m, a shift by one pixel (0.17 μ m) per every section (1 μ m) corresponds to a tilt of 9.6°. Therefore the stereo pair is separated by a tilt of 19.2°. To remove out-of-focus information from the projected images, a three-dimensional optical transfer function was experimentally determined for the objective lens under the conditions used for the data collection of nuclei by imaging very small fluorescent beads (0.1 μ m in diameter) as described previously⁵. The projected stereo images were computationally processed to remove out-of-focus information by a constrained iterative deconvolution method using the Z^* =0 plane of the three-dimensional optical transfer function⁴. This approach, based



on the Fourier projection theorem, can provide nearly perfect signal recovery using data collection conditions that cause minimum exposure to the specimen. Although not precisely correct for the sliding stereo projections used here, the approximation is suitable for the limited spatial regions examined. For rapid movie display, the entire set of time-lapse stereo images was loaded into the Parallax display system with a 12 Mbyte image memory before the display. This permits a pair of the stereo projection images to be displayed side-by-side and rapidly updated to form a time-lapse stereographic movie.

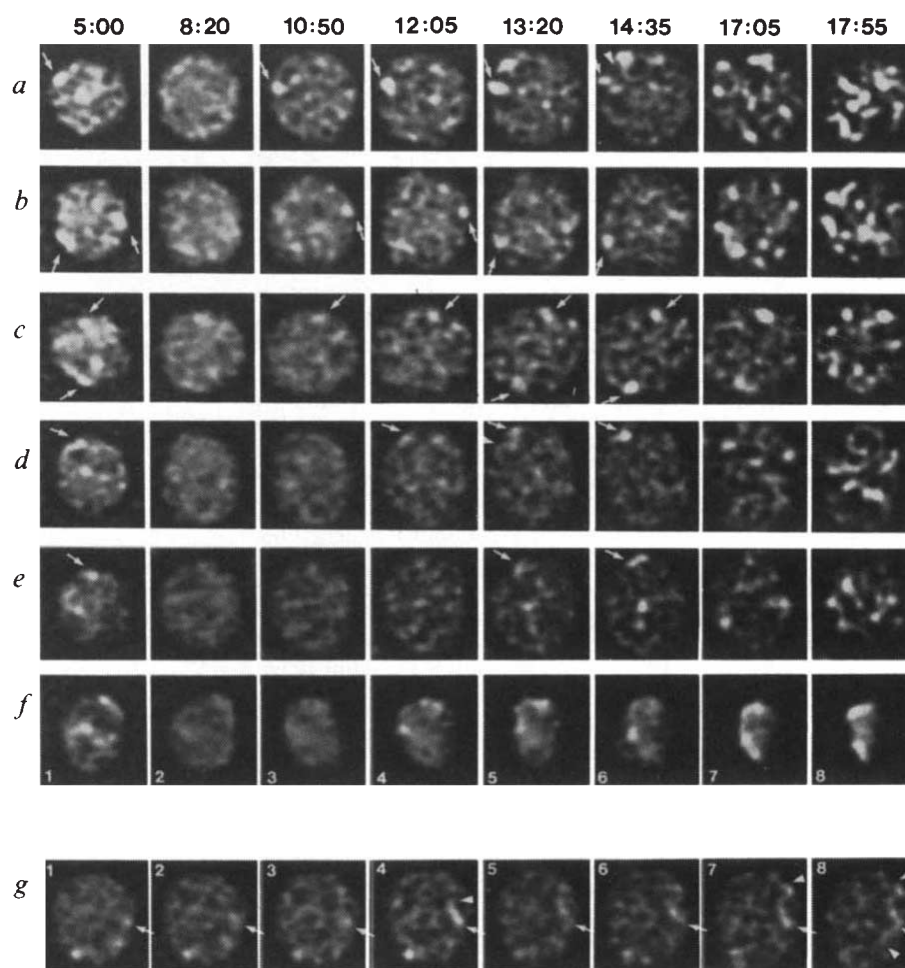


FIG. 2 Chromosome decondensation and condensation. Chromosome decondensation and condensation is shown as a time course. Only one view of the stereo movie is shown for each time point. A damaged nucleus is shown in *f*. Chromosome fibres growing in a single nucleus shown in *g* at every 25 s from 10 min 50 s to 13 min 45 s. The numbers at the top of each column or frame represent time in minutes:seconds as in Fig. 1. Nucleation sites are shown by arrows and fibres by arrowheads. Scale bar represents 5 μ m.

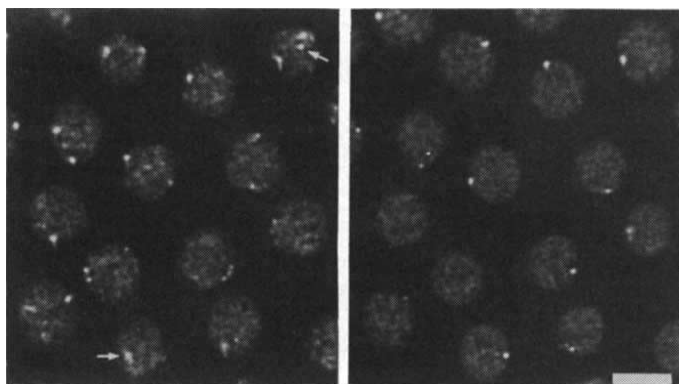


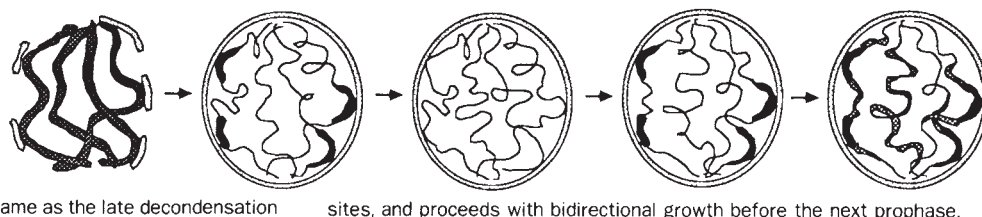
FIG. 3 Sites of condensed chromatin in fixed nuclei. Two optical section images are shown, separated by 2 μ m, from a high-resolution three-dimensional data set of a fixed embryo. Because of the curvature of the embryo surface, the focal plane intersects nuclei seen in the upper and lower portions of the field at higher positions than those located in the centre of the field. All of the nuclei shown in the left panel are visualized above the nuclear midline, whereas those in the right panel are all at, or below, the midline. Arrows indicate apparent intranuclear sites above the midline. Scale bar represents 5 μ m.

METHODS. *Drosophila* embryos were fixed with 3.7% formaldehyde in heptane/buffer A (15 mM PIPES, pH 7.0, 80 mM KCl, 20 mM NaCl, 0.5 mM EGTA, 2 mM EDTA, 0.5 mM spermidine, 0.2 mM spermine, 0.1% 2-mercaptoethanol). The chorion and vitelline membrane were removed as described previously¹⁷. The embryos were stained with 0.1 μ g ml⁻¹ DAPI. Optical section images were collected on the CCD at focus step of 0.2 μ m using an Olympus oil-immersion objective lens (60 $\times$ /NA1.4) and computationally processed to remove the out-of-focus information as described⁴.

appear over a 5-minute interval (Fig. 2*a-e*, columns 3–6). Significantly, the relative positions of these early condensation sites were similar to those of the late decondensation sites. Examples of the congruity as shown in Fig. 2*a-e* (compare column 1 with columns 3–6). In addition, newly condensed chromosome 'fibres' can be seen elongating from these sites at the rate of about 1 μ m per min (Fig. 2*a, d* and *g*). In one particularly clear case (Fig. 2*g*), the elongation appears to be bidirectional. As the condensation sites were found near the nuclear midline, they are unlikely to be either centromeric or telomeric sites, which are known to be localized to the apical and basal portions of the nucleus respectively⁹. Towards prophase (Fig. 2*a-e*, columns 7, 8), the chromosomes are seen to condense further, forming more contiguous structures. By the time condensation has proceeded sufficiently to identify individual chromosomes, the sites have detached and moved from the nuclear envelope (compare columns 6 and 7 in Fig. 2*a-e*), making it impossible to assign these sites, as yet, to specific chromosomes.

Similar chromosomal bright spots were observed in higher-resolution optical section data of nuclear-cycle-13 embryos that had been fixed and stained with the DNA-specific fluorescent dye, 4',6-diamidino-2-phenylindole (DAPI), as shown in Fig. 3. Below the nuclear midline, chromosome condensation sites were exclusively observed at the nuclear envelope. By contrast, sites were seen both internally and at the periphery above the midline. It seems that the internal bright spots are associated with, and likely to be a part of, the centromeric heterochromatin. Although such peripheral bright spots had been repeatedly observed in fixed preparations in our laboratory, it is only as a result of continuous observation of chromosome behaviour in individual nuclei that we appreciated the significance of these chromosomal regions.

FIG. 4 Schematic diagram of the events of chromosome decondensation and condensation. After the nuclear envelope is reformed on telophase chromosomes, decondensation proceeds sequentially, terminating at specific sites at or near the nuclear envelope. Chromosome condensation starts again at discrete sites, possibly the same as the late decondensation



sites, and proceeds with bidirectional growth before the next prophase.

Using the time-lapse data set, we also have been able to analyse the behaviour of a damaged nucleus (indicated by asterisks in Fig. 1). A chromosomal cross-bridge, the consequence of naturally occurring mitotic failure or photo-induced crosslinking of chromatin, is seen to be extended between daughter nuclei during anaphase. This anaphase bridge was eventually resolved, probably by chromosomal breakage. The resultant damaged nuclei proceeded through chromosome decondensation and condensation with about the same timing as neighboring undamaged nuclei (Fig. 2f; see also Fig. 1), but failed to divide at the next nuclear division cycle (Fig. 1). The damaged nuclei then migrated to the embryo interior while normal nuclei divided and remained at the embryo surface. The loss of a small number of isolated nuclei at the syncytial blastoderm stage apparently does not affect further embryonic development³.

These results suggest that sites on chromosomes or on the nuclear envelope function as nucleation points of condensation and decondensation of chromosomes (Fig. 4). Furthermore, the congruity of those sites at telophase and the next prophase suggests that nuclear structures are, in part, preserved during interphase. Previously, we have shown that in *D. melanogaster* polytene interphase chromosomes, there are 15–20 chromosomal sites that are involved in attachment to the nuclear envelope^{10–12}. The analysis of chromosomal rearrangements demonstrated that the attachments are chromosomal site-specific, and nuclear-envelope non-specific¹³. As yet, there is no direct evidence for the correlation of these sequences with the condensation initiation loci observed in this report. As many features of chromosome topology are clearly below the resolution of light microscopy, we cannot eliminate the possibility that some limited chromosome condensation might take place at sites other than on the nuclear envelope. Precise determination of the spatial distribution of sites will require analyses of higher resolution.

The real-time, three-dimensional analysis used in this work represents a compromise between temporal and spatial resolution. To speed up the image acquisition rate, data were collected at a spatial resolution lower than we normally use when examining fixed specimens (pixel sizes: 0.17 μm versus 0.06 μm in the microscopic field plane and 1 μm versus 0.2 μm in the focus direction). Nevertheless, examining whole cells in the living state in three dimensions can provide unique insights into cellular structure and function not possible with other structural methods. Improvements in data collection software and the use of a faster CCD readout rate will decrease the time required to record and store an image by fivefold. Storing only final stereo pairs instead of individual section images should provide a further twofold improvement in speed.

To understand cellular events, it is important to visualize macromolecular interactions within the three-dimensional context of the cell as a function of time. Other 'four-dimensional' imaging schemes using polarization or differential interference contrast microscopy have been devised¹⁴, but the molecular specificity of imaging in fluorescence microscopy makes it the technique of choice for our studies of interphase nuclei. In addition, by the use of rapid wavelength-switching, it should be possible to extend the 'four-dimensional' technique described here to the simultaneous analyses of multiple macromolecular components in individual living cells^{15,16}. We anticipate that the

approach can be extended to a wide range of dynamic biological phenomena. □

Received 2 May; accepted 4 October 1989.

1. Comings, D. E. & Okada, T. A. *Expl. Cell Res.* **63**, 471–473 (1970).
2. Robbins, E., Pederson, T. & Klein, P. *J. Cell Biol.* **44**, 400–416 (1970).
3. Minden, J. S., Agard, D. A., Sedat, J. W. & Alberts, B. *J. Cell Biol.* **109**, 505–516 (1989).
4. Agard, D. A., Hiraoka, Y., Shaw, P. & Sedat, J. W. *Meth. Cell Biol.* **30**, 353–377 (1989).
5. Hiraoka, Y., Sedat, J. W. & Agard, D. A. *Science* **238**, 36–41 (1987).
6. Aikens, R. S., Agard, D. A. & Sedat, J. W. *Meth. Cell Biol.* **29**, 291–313 (1989).
7. Zalokar, M. & Erk, I. *J. Microsc. Biol. Cell.* **25**, 97–106 (1976).
8. Foe, V. E. & Alberts, B. M. *J. Cell Biol.* **61**, 31–70 (1983).
9. Foe, V. E. & Alberts, B. M. *J. Cell Biol.* **100**, 1623–1636 (1985).
10. Mathog, D., Hochstrasser, M., Gruenbaum, Y., Saumweber, H. & Sedat, J. W. *Nature* **308**, 414–421 (1984).
11. Hochstrasser, M., Mathog, D., Gruenbaum, Y., Saumweber, H. & Sedat, J. W. *J. Cell Biol.* **102**, 112–123 (1986).
12. Hochstrasser, M. & Sedat, J. W. *J. Cell Biol.* **104**, 1471–1483 (1987).
13. Mathog, D. & Sedat, J. W. *Genetics* **121**, 293–311 (1989).
14. Inoue, S. & Inoue, T. D. *Ann. N.Y. Acad. Sci.* **483**, 392–404 (1987).
15. Taylor, D. L., Amato, P. A., McNeil, P. L., Luby-Pheips, K. & Tanasugarn, L. in *Applications of Fluorescence in the Biomedical Sciences* (eds Taylor, D. L., Waggoner, A. S., Murphy, R. F., Lanni, F. & Birge, R. R.) 347–376 (Liss, New York, 1986).
16. Waggoner, A. et al. *Meth. Cell Biol.* **30**, 449–478 (1989).
17. Mitchison, T. J. & Sedat, J. W. *Dev. Biol.* **99**, 261–264 (1983).

ACKNOWLEDGEMENTS. We thank Hans Chen and Mel Jones for their crucial contributions to software and hardware development and Elizabeth Blackburn, Mary Rykowski, Andrew Belmont and Tokuko Haraguchi for critical reading of the manuscripts. This work was supported by the National Institutes of Health and by the Howard Hughes Medical Institute (J.W.S. and D.A.A.). D.A.A. is also a National Science Foundation Presidential Young Investigator. J.S.M. was supported by a fellowship from the Helen Hay Whitney Foundation and Y.H. a Damon Runyon-Walter Winchell Cancer Research Fund Fellowship.

Capping and α -helix stability

Luis Serrano & Alan R. Fersht

MRC Unit for Protein Function and Design, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

THE first and last four residues of α -helices differ from the rest by not being able to make the intrahelical hydrogen bonds between the backbone $>\text{C}=\text{O}$ groups of one turn and the $>\text{NH}$ groups of the next¹. Physico-chemical arguments¹ and statistical analysis² suggest that there is a preference for certain residues at the C and N termini (the C- and N-caps)² that can fulfil the hydrogen bonding requirements. We have tested this hypothesis by constructing a series of mutations in the two N-caps of barnase (*Bacillus amyloliquefaciens* ribonuclease, positions Thr 6 and Thr 26) and determining the change in their stability. The N-cap is found to stabilize the protein by up to $\sim 2.5 \text{ kcal mol}^{-1}$. The presence of a negative charge of the N-cap adds some $1.6 \text{ kcal mol}^{-1}$ of stabilization energy because of the interaction with the macroscopic electrostatic dipole of the helix.

The α -helix is a major component of protein secondary structure. Cogent arguments have been made from simple physical chemistry¹, statistical survey² and qualitative experiments on model systems^{3,4} to identify the structural features that influence helix stability. Quantitative data on the energetics and importance of these interactions, however, have been unavailable until very recently. The small ribonuclease from *B. amyloliquefaciens*, barnase, has proved to be an excellent system for determining these with high precision from measurements on unfolding of

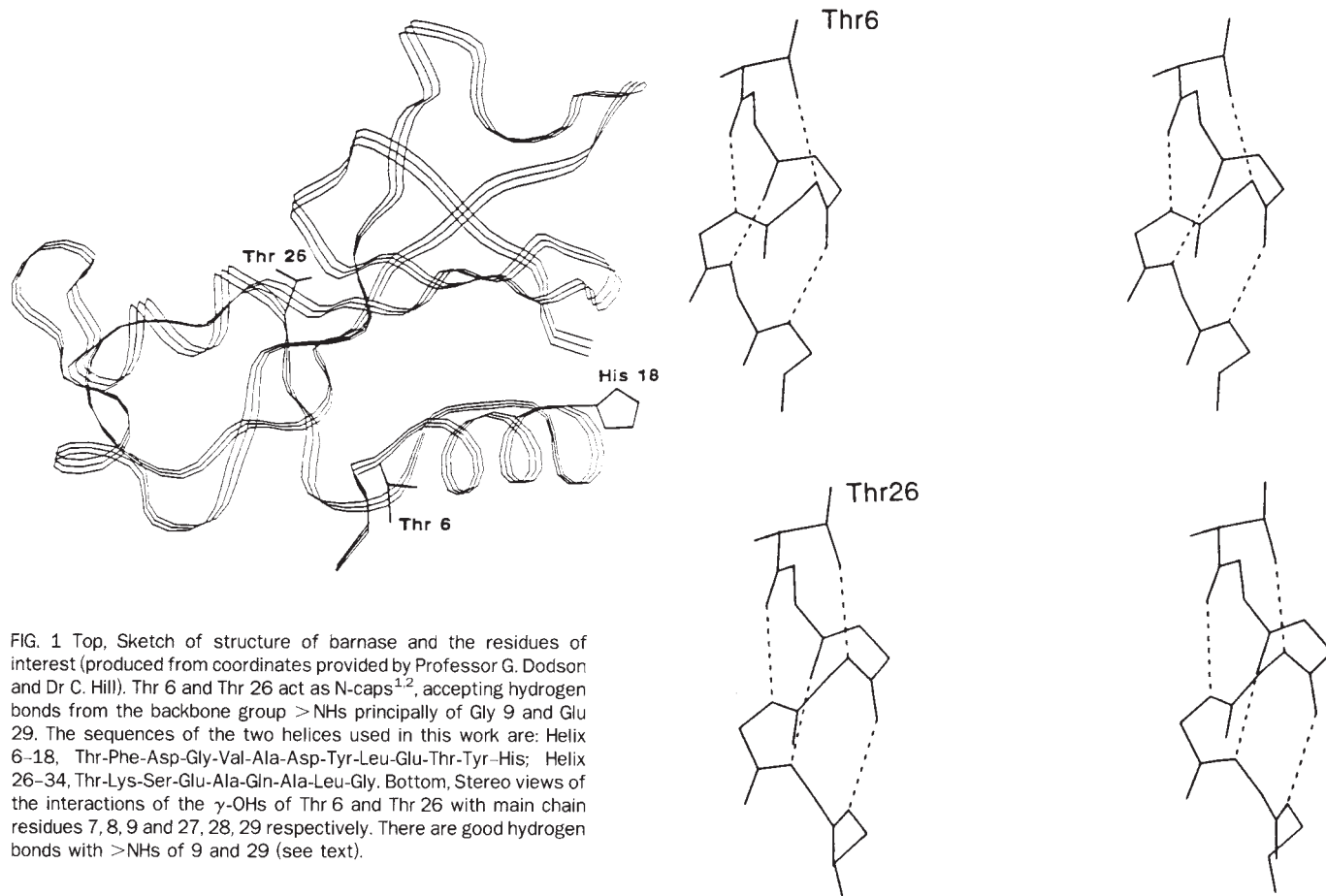


FIG. 1 Top, Sketch of structure of barnase and the residues of interest (produced from coordinates provided by Professor G. Dodson and Dr C. Hill). Thr 6 and Thr 26 act as N-caps^{1,2}, accepting hydrogen bonds from the backbone group $>NH$ s principally of Gly 9 and Glu 29. The sequences of the two helices used in this work are: Helix 6-18, Thr-Phe-Asp-Gly-Val-Ala-Asp-Tyr-Leu-Glu-Thr-Tyr-His; Helix 26-34, Thr-Lys-Ser-Glu-Ala-Gln-Ala-Leu-Gly. Bottom, Stereo views of the interactions of the γ -OHs of Thr 6 and Thr 26 with main chain residues 7, 8, 9 and 27, 28, 29 respectively. There are good hydrogen bonds with $>NH$ s of 9 and 29 (see text).

wild-type and mutant proteins⁵⁻⁸. Barnase contains two α -helices⁹, one from residues 6 to 18 and the other from 26 to 34 (Fig. 1). Both α -helices begin with a threonine residue as N-cap (the first residue in the helix², namely Thr 6 and Thr 26, Fig. 1), but the overall composition of the rest of each helix is quite different. The -OH groups of Thr 6 and Thr 26 are positioned so that the oxygen atom may act as a hydrogen bond acceptor from the backbone $>NH$ groups of residues 7 or 27, 8 or 28 and 9 or 29 (N1, N2 and N3 in the Richardson² nomenclature). The geometry for hydrogen bonding from the -OH of residues 6 and 26 is good with respect to residues 9 and 29, fair for 8 and 28 and poor for 7 to 27 (Table 1), which is usual with N-caps². The C_γ and C_β of both threonines participate in further interactions in the backbone (Table 1). Here we use inferences from recent surveys of helix structure^{1,2} to design changes in both barnase helices to give precise information on how hydrogen bonding, and electrostatic¹⁰ and other interactions at the N-terminus of a helix affect its stability.

The most common residues found at the N-cap position of helices are: Ser, Asn, Gly, Asp and Thr². The residues Ala, Leu, Val, Ile, Trp, Arg, Gln and Glu are found rarely². We substituted Thr with the following residues: with Gly, Ala and Val to measure the effect of losing hydrogen bonds from the γ -OH to the $>NH$ groups of the helix and to assess the influence of a different size of side chain; with Ser to measure the loss of hydrophobic interactions from the γ -methyl; with Asn and Gln to determine the effects of substituting the carboxamide oxygen of each of the γ -OH as a hydrogen bond acceptor, and with Asp and Glu to establish the effects of a charged hydrogen bond acceptor. Two mutants, Thr 26 $\rightarrow$ Asp and Thr 6 $\rightarrow$ Val, could not be expressed, despite numerous attempts using this expression system¹¹.

Mutation of either Thr 6 or Thr 26 at the N-caps of each helix in barnase gives quite similar changes in stability and consistent

trends (Table 2). Considering that the two helices have different compositions and environments, the results are not specific to a particular helix and so presumably report on some general features. A fine-structure analysis can be made by comparing pairs of mutants that differ by just a small change.

Thr $\rightarrow$ Ser. Mutation of Thr 6 $\rightarrow$ Ser causes the loss of 0.22 kcal mol⁻¹ and Thr 26 $\rightarrow$ Ser, 0.56 kcal mol⁻¹ of stabilization energy (Table 2). The C_γ of Thr 6 and 26 make van der Waals' contacts with the protein (Table 1), and so this contribution to the stabilization is lost. It should be noted that none of these contacts is within the helix. Thus, the advantage of having a threonine rather than a serine at the N-cap seems to depend mainly on the existence of tertiary interactions elsewhere in the structure, and not on interactions within the helix itself. These interactions will vary from protein to protein.

Ser $\rightarrow$ Gly, Ala or Val. Naïve reasoning would suggest that mutation of Ser $\rightarrow$ Ala might give an empirical measurement of energy change due to the loss of the hydrogen bonds from the -OH to the backbone $>NH$ groups, and Ala $\rightarrow$ Gly, a measure of the loss of the van der Waals' energy from the β -methyl group. Mutation of Ser 6 $\rightarrow$ Ala loses 2.3 kcal mol⁻¹ or 1.6 kcal mol⁻¹ for replacement of Ser 26 (Table 2). But, mutation of Ser $\rightarrow$ Gly loses only 1 kcal mol⁻¹. Thus, Gly is more stable than Ala at the N-cap by 0.5-1 kcal mol⁻¹. In general, Ala is energetically preferred to Gly in the helix, except at the N-cap¹², either because of the greater entropic freedom of glycine-containing peptides in the denatured state¹³, or because of hydrophobic interactions of the alanine β -methyl group in the helix. Why then is Gly preferred to Ala at the N-cap? We suggest that one contributing factor is that the β -methyl group of Ala sterically interferes with solvation of the $>NH$ groups at the N-terminus of the helix—Gly allows more space for optimal hydrogen bonding by water. Consistent with this notion is that Thr 26 $\rightarrow$ Val is the most destabilizing mutation, despite the retention of the

γ -methyl group with its potential for hydrophobic interactions worth ~ 0.6 kcal mol $^{-1}$. The large side chain of Val makes solvation of the $>NH$ groups by water difficult. These results parallel statistical data on the frequency of residues at the N-cap (Table 3). There is evidence elsewhere that an uncharged hydrogen bond contributes some 0.5–1.5 kcal mol of stabilization energy to a complex¹⁴. The apparent binding energy of the N-cap hydrogen bond from Ser to the $>NH$ groups is in this region.

Thr $\rightarrow$ Asn, Gln. The previous mutations are relatively easy to interpret in the absence of detailed crystal structures of the mutants because interactions were deleted without introducing new ones, and so the necessary structural information is present in the crystal structure of the wild type. Substitution with Asn, Asp (or Glu or Gln), adds the possibility of further interactions that can obscure simple interpretation. Statistically, Asn is one of the most favourable residues at the N-cap of helices². The carbonyl oxygen of the side chain can act as a hydrogen bond acceptor of the backbone $>NH$ groups. On the other hand, Gln and Glu are rarely found at the N-cap as their side chains cannot make good hydrogen bonds to the backbone $>NH$ groups². Mutation of Thr $\rightarrow$ Gln in positions 6 to 26 results in destabilizing the protein by 1.9 and 1.7 kcal mol $^{-1}$ respectively. But Thr $\rightarrow$ Asn also destabilizes both helices by 1.3 kcal mol $^{-1}$. The reason for this is not obvious, because Asn can be built into the structure with its carbonyl oxygen placed to accept hydrogen bonds from

the backbone $>NH$ groups and its own $-NH_2$ group exposed to water. Perhaps Asn is not inherently as good an N-cap as Ser or Thr, but may be better when the $-NH_2$ group can hydrogen bond elsewhere in the protein. There should be more protein structure surveys undertaken to look for other factors. Whatever the explanation, Asn is not necessarily a good choice for the N-cap of an helix.

Thr $\rightarrow$ Asp, Glu. Mutation of Thr 6 $\rightarrow$ Asp stabilizes the protein by 0.1 kcal mol $^{-1}$, but Thr 6 $\rightarrow$ Glu destabilizes it by only 0.3 kcal mol $^{-1}$ and Thr 26 $\rightarrow$ Glu destabilizes it by less than 0.1 kcal mol $^{-1}$, despite the low frequency of Glu at the N-cap². The α -helix possesses a macroscopic dipole with its positive pole at the N-terminus¹⁰. Recent studies have suggested values of about 1.4–2.1 kcal mol $^{-1}$ for the interaction energy of a single charge with the dipole of a helix^{6,15}. A charge–dipole electrostatic interaction is the probable reason for the stability of the mutants with Asp and Glu. Comparison of Asp with Asn and Glu with Gln gives another estimate of the importance of the charge–dipole interaction: Asp 6 $\rightarrow$ Asn gives 1.4 kcal mol $^{-1}$; Glu 6 $\rightarrow$ Gln, 1.6 kcal mol $^{-1}$; and Glu 26 $\rightarrow$ Gln, 1.6 kcal mol $^{-1}$. These results are strikingly similar to those of the earlier studies, despite

TABLE 1 Contact distances (< 5 Å) from Thr 6 and Thr 26 to other residues*

Atom	Neighbouring atom	Distance (Å)
Thr 6	O γ	N (Phe 7) 3.63, 3.53, 3.80 †
		N (Asp 8) 4.17, 4.06, 4.00 †
		N (Gly 9) 3.57, 3.14, 3.05 †
	C γ	O δ (Asn 5) 3.83
		C (Asn 5) 3.81
		O (Asn 5) 4.20
	C β	O δ (Asn 77) 3.98
		N (Phe 7) 3.05
		N (Asp 8) 4.55
		N (Gly 9) 4.73
Thr 26	O γ	C (Asn 5) 3.64
		O (Asn 5) 4.15
		N (Lys 27) 3.37, 3.34, 3.67 †
		N (Ser 28) 3.54, 3.38, 3.77 †
		N (Glu 29) 3.35, 2.94, 3.23 †
	C γ	O δ_1 (Asp 54) 4.20
		O σ_2 (Asp 54) 4.77
		C (Ile 25) 4.04
		O (Gly 52) 3.52
		C (Gly 53) 4.48
		O (Gly 53) 4.47
		O δ_1 (Asp 54) 4.07
		O δ_2 (Asp 54) 4.66
		N (Lys 27) 3.08
		N (Ser 28) 4.30
	C β	N (Glu 29) 4.66
		O (Ile 25) 4.34
		C (Ile 25) 3.70
		O (Gly 52) 4.17
		C γ (Asp 54) 3.88
		O δ_2 (Asp 54) 3.74

* Contact distances from Thr 6 and Thr 26 to other residues were obtained from X-ray coordinates (2 Å resolution), kindly provided by Professor G. Dodson and Dr C. Hill.

† There are three molecules in the unit cell, and these are the individual values for each molecule. The accuracy for bond distances at this resolution is about ± 0.2 Å.

TABLE 2 Change in free energy of unfolding ($\Delta\Delta G_U$) on mutation of N-caps (Thr 6 and Thr 26)*

Mutation	$\Delta\Delta G_U$ (kcal mol $^{-1}$)	
	Position 6	Position 26
Thr $\rightarrow$ Ser	0.22	0.56
Thr $\rightarrow$ Val		2.31
Thr $\rightarrow$ Ala	2.53	2.11
Thr $\rightarrow$ Gly	1.34	1.58
Thr $\rightarrow$ Asn	1.27	1.29
Thr $\rightarrow$ Asp	−0.11	
Thr $\rightarrow$ Gln	1.87	1.72
Thr $\rightarrow$ Glu	0.27	0.05
Asp $\rightarrow$ Asn	1.38	
Glu $\rightarrow$ Gln	1.60	1.67
Asp $\rightarrow$ Glu	0.38	
Asn $\rightarrow$ Gln	0.60	0.43
Ser $\rightarrow$ Ala	2.31	1.55
Ser $\rightarrow$ Gly	1.12	1.02
Gly $\rightarrow$ Ala	1.19	0.53

* $\Delta\Delta G_U$ is the free energy of unfolding of wild-type enzyme minus that of mutant. The free energies of unfolding of wild-type protein and mutants were analysed by urea denaturation monitored by fluorescence^{5–7}. This analysis is based on the equation $\Delta G_U = \Delta G_U^{H_2O} - m[Urea]$, where ΔG_U is the free energy of unfolding in urea solution, $\Delta G_U^{H_2O}$ the free energy of unfolding in water, and m is a constant. We find that there is little difference in calculating $\Delta\Delta G_U$ directly by comparing ΔG_U for wild type and mutant at the same concentration of urea⁵ or by using the simpler procedure⁷ of $\Delta\Delta G_U = \langle m \rangle (\Delta[U]_{1/2})$, where $\Delta[U]_{1/2}$ is the concentration of urea at which wild type is 50% denatured, minus the concentration of urea at which the mutant is 50% denatured. The average value of m for wild type and mutant is denoted by $\langle m \rangle$. We find that m does not deviate by more than 10% of a mean value which is 2.23 (± 0.02 standard error) kcal mol $^{-2}$ for all mutants. $[U]_{1/2}$ never differs by more than 1% ($< \pm 0.04$ M) for experiments performed over a period of time and is generally reproducible to ± 0.01 M with identical batches of urea. The data are probably reliable to ± 0.02 kcal mol $^{-1}$. As the free energy of unfolding determined by either method is essentially the same, we have used the second method to obtain the values of the free energy of unfolding of all the proteins.

Proteins were prepared by an improved procedure. Wild-type enzyme was secreted from *E. coli* BL21DE3 pLysS (ref. 16) harbouring the barnase gene construction¹² on the pTZ18u plasmid (Pharmacia). Cells were grown in low-phosphate medium¹⁷ and the expression of barnase was induced by the addition of 0.5 mM isopropyl- β -D-thiogalactoside following the conditions described by Studier *et al.*¹⁶. After further growth for 2.5 h, the protein was purified as before¹⁸. For site-directed mutagenesis we used the method of Eckstein *et al.*¹⁹. Mutants were identified either by direct sequencing of single-stranded DNA or after previous screening by the hybridization method²⁰.

TABLE 3 Comparison of statistical and energetic surveys of N-cap stabilities

Preferred residue at N-cap	Stabilization energy relative to Thr (kcal mol ⁻¹)*	Frequency at N-cap†	Relative value‡
Asp	-0.1	27	2.1
Thr	0.0	21	1.6
Glu	0.2	5	0.4
Ser	0.4	34	2.3
Asn	1.3	34	3.5
Gly	1.5	33	1.8
Gln	1.8	3	0.4
Ala	2.3	10	0.5
Val	2.3	1	0.1

* Average value at positions 6 and 26.

† Frequency found in survey in ref. 2.

‡ Frequency at N-cap normalized for frequency of occurrence in proteins in general².

there being an unfavourable electrostatic interaction at position 6 with the negative charge on Asp 8. As so many different substitutions with different hydrogen bonding patterns and at different ends of the helix give similar values, the estimate of 1.4–2.1 kcal mol⁻¹ for the helix–charge interaction is valid.

There is a very rough parallel between the energetics of N-cap formation and statistical surveys of the frequencies found in practice (Table 3). Notable exceptions, however, are Asn and Glu. Our data indicate that Glu is quite acceptable at the N-cap, being only marginally less stable than Thr. Asn is statistically over-represented, given the data in this study.

In conclusion, the N-cap is important to stability and substitution of selected residues can cause an energy change of over 2 kcal mol⁻¹. It is clear, however, that when designing a helix, the choice of residue depends on the context of the overall tertiary structure—that is, on interactions with residues removed in sequence but stereochemically close to the helix. Even so, it seems that some residues are particularly favourable. For example, Ser is inferior to Thr only when hydrophobic interactions are favourable for the γ -methyl group. For barnase, the Ser mutants are only 0.2 to 0.6 kcal mol⁻¹ less stable, although the methyl group can be worth up to 1.5 kcal mol⁻¹ when packing is particularly good^{5,7}. In other circumstances, steric exclusion of the γ -methyl group would prohibit the use of Thr. Asp is clearly the best for an exposed N-terminus that has no adjacent negative charges. □

X-ray analysis of HIV-1 proteinase at 2.7 Å resolution confirms structural homology among retroviral enzymes

Risto Lapatto, Tom Blundell*, Andrew Hemmings, John Overington, Andrew Wilderspin, Stephen Wood, James R. Merson†, Peter J. Whittle‡, Dennis E. Danley‡, Kieran F. Geoghegan‡, Steven J. Hawrylik‡, S. Edward Lee‡, Kathryn G. Scheld‡ & Peter M. Hobart‡

Laboratory of Molecular Biology, Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, UK

† Pfizer Central Research, Ramsgate Road, Sandwich, Kent CT13 9NJ, UK

‡ Pfizer Central Research, Eastern Point Road, Groton, Connecticut 06340, USA

KNOWLEDGE of the tertiary structure of the proteinase from human immunodeficiency virus HIV-1 is important to the design of inhibitors that might possess antiviral activity and thus be useful in the treatment of AIDS¹. The conserved Asp–Thr/Ser–Gly sequence in retroviral proteinases² suggests that they exist as dimers similar to the ancestor proposed for the pepsins^{3–5}. Although this has been confirmed by X-ray analyses of Rous sarcoma virus and HIV-1 proteinases^{6,7}, these structures have overall folds that are similar to each other only where they are also similar to the pepsins⁸. We now report a further X-ray analysis of a recombinant HIV-1 proteinase at 2.7 Å resolution. The polypeptide chain adopts a fold in which the N- and C-terminal strands are organized together in a four-stranded β -sheet. A helix precedes the single C-terminal strand, as in the Rous sarcoma virus proteinase⁶ and also in a synthetic HIV-1 proteinase, in which the cysteines have been replaced by α -aminobutyric acid⁹. The structure reported here provides an explanation for the amino acid invariance amongst retroviral proteinases, but differs from that reported earlier⁷ in some residues that are candidates for substrate interactions at P₃, and in the mode of intramolecular cleavage during processing of the polyprotein.

To obtain HIV-1 proteinase for crystallization, a recombinant fragment of the viral *pol* gene was expressed in *E. coli* as described in the legend to Fig. 1. Crystals, prepared using the published protocol¹⁰, grew at 4 °C in five days. X-ray data were collected from a tetragonal bipyramidal crystal to 2.7 Å resolution using a FAST area detector. Earlier, we constructed models using Composer⁵, initially from the pepsins and later from the Rous sarcoma virus (RSV) structure. From these we constructed a model for use in molecular replacement that did not include the N- or C-terminal strands, the helix or the flap, the positions of which were uncertain. A series of difference Fourier analyses, modelling, and refinement steps gave first the region 47 to 54 (the flap), then the N- and C-termini, and finally the turn after the N-terminal strand (residues 5 to 11) and the α -helix (residues 87 to 93). In each case the polypeptide could be followed in the difference electron density (calculated on the basis of a model in which they had not previously been included) without a break and with clear positions for the side chains (Fig. 1). Thus we achieved an unequivocal tracing of all 99 residues of the polypeptide chain. Further details of the analysis and refinement are given in the legend to Fig. 1. The final geometry has a 0.019 Å r.m.s. deviation from idealized geometry and the present agreement value is 18.9%.

Received 23 June; accepted 6 October 1989.

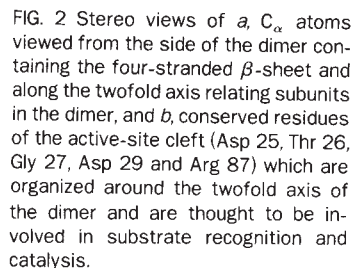
- Presta, L. G. & Rose, G. D. *Science* **240**, 1632–1641 (1988).
- Richardson, J. S. & Richardson, D. C. *Science* **240**, 1648–1652 (1988).
- Kim, P. S. & Baldwin, R. L. *Nature* **307**, 329–334 (1984).
- Shoemaker, K., Kim, P. S., York, E. J., Stewart, J. M. & Baldwin, R. L. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2349–2353 (1985); *Nature* **326**, 563–567 (1987).
- Kellis, J. T. Jr, Nyberg, K., Sali, D. & Fersht, A. R. *Nature* **333**, 784–786 (1988).
- Sali, D., Bycroft, M. & Fersht, A. R. *Nature* **335**, 741–743 (1988).
- Kellis, J. T. Jr, Nyberg, K. & Fersht, A. R. *Biochemistry* **28**, 4914–4922 (1989).
- Matouschek, A., Kellis, J. T., Serrano, L. & Fersht, A. R. *Nature* **340**, 122–126 (1989).
- Mauguen, Y. *et al.* *Nature* **297**, 162–164 (1982).
- Hol, W. G. J. *Prog. Biophys. molec. Biol.* **45**, 149–195 (1985).
- Paddon, C. J. & Hartley, R. W. *Gene* **53**, 11–19 (1987).
- Strehlow, G. H. & Baldwin, L. R. *Biochemistry* **28**, 2130–2133 (1989).
- Matthews, B. W., Nicholson, H. & Beckett, W. J. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6663–6667 (1987).
- Fersht, A. R. *Trends Biochem. Sci.* **12**, 301–304 (1987).
- Nicholson, H., Beckett, W. J. & Matthews, B. W. *Nature* **336**, 651–656 (1988).
- Studier, F. W., Rosenberg, A. H. & Dunn, J. J. *Meth. Enzym.* (in the press).
- Serpesu, E. H., Shortle, D. & Mildvan, A. S. *Biochemistry* **25**, 68–87 (1986).
- Mossakowska, D. N., Nyberg, K. & Fersht, A. R. *Biochemistry* **28**, 3843–3850 (1989).
- Sayers, J. R. & Eckstein, F. In *Genetic Engineering: Principles and Methods* Vol. 10 (ed. Setlow, J. K.) 109 (Plenum, New York and London, 1988).
- Carter, P. J., Winter, G., Wilkinson, A. J. & Fersht, A. R. *Cell* **38**, 835–840 (1984).

ACKNOWLEDGEMENTS. We thank Professor Guy Dodson and Dr Chris Hill for providing coordinates from unpublished high resolution data on barnase. L.S. is an EMBO fellow.

* To whom correspondence should be addressed.

METHODS. Enzyme for crystallization was expressed in *E. coli* as follows. The bacterial expression plasmid pHVexpol5 was constructed by subcloning the 5.0 kilobase *Bgl*II fragment from the pHXbc2 clone of the viral genome¹⁸ into the *Bgl*II-cleaved derivative of pPFZ-R2 (ref. 19). This gave a gene controlled by the *trp* promoter and encoding a fusion polypeptide comprising an initiator methionine, a 7-residue junction sequence Ala-Glu-Ile-Thr-Arg-Ile-Glu-, and a fragment of the *pol* gene product which had a relative molecular mass of 115,000 (115K) which encompassed the proteinase, reverse transcriptase and integrase domains. *E. coli* MM294 cells harbouring pHVexpol5 were shown by western blotting and immunodetection to produce appropriately processed proteinase (11K). HIV-1 proteinase was purified from the soluble fraction of cell lysates (D.E.D. *et al.*, manuscript in preparation) to give a single band on Coomassie blue-stained SDS-polyacrylamide gels. Automated Edman degradation gave an N-terminal sequence of Pro-Gln-Ile-Thr-Leu-, consistent with the expected result of self-processing of the *pol* gene product. The purified proteinase recovered per litre of bacterial culture was 0.25 mg. The specific activity of the purified enzyme was measured using a 17-residue synthetic peptide substrate (RRVTNSATIMMQRGNFR (one-letter code); J. Schneider, personal communication), and was estimated at 2.4 μ mol substrate cleaved per min per mg of proteinase at pH 5.5, 23 °C and 0.2 mM substrate. Details of the expression studies, cell growth and protein

purification will be presented elsewhere (D.E.D. *et al.*, manuscript in preparation). Crystals, isomorphous with those reported earlier⁷ of space group $P4_12_12$, were used for FAST area detector data collection, giving 10,000 reflections over a period of two days. A merging R -value was 0.11 leading to over 2,200 reflections with $I > 2.0\sigma(I)$, with a maximum resolution of 2.7 Å. Refinement of the initial model using rigid group refinement was achieved using RESTRAIN²⁰ using the CRAY XMP/28 at University of London Computing Centre. In the final cycles, we used individual thermal parameters but tight geometric restraints. Difference Fouriers calculated with coefficients $2F_{\text{obs}} - F_{\text{calc}}$ and $F_{\text{obs}} - F_{\text{calc}}$ were displayed simultaneously using FRODO²¹ on an Evans and Sutherland PS390. The model was extended in steps to include the N- and C-termini and the helix and at each stage refined using restrained least squares.



b

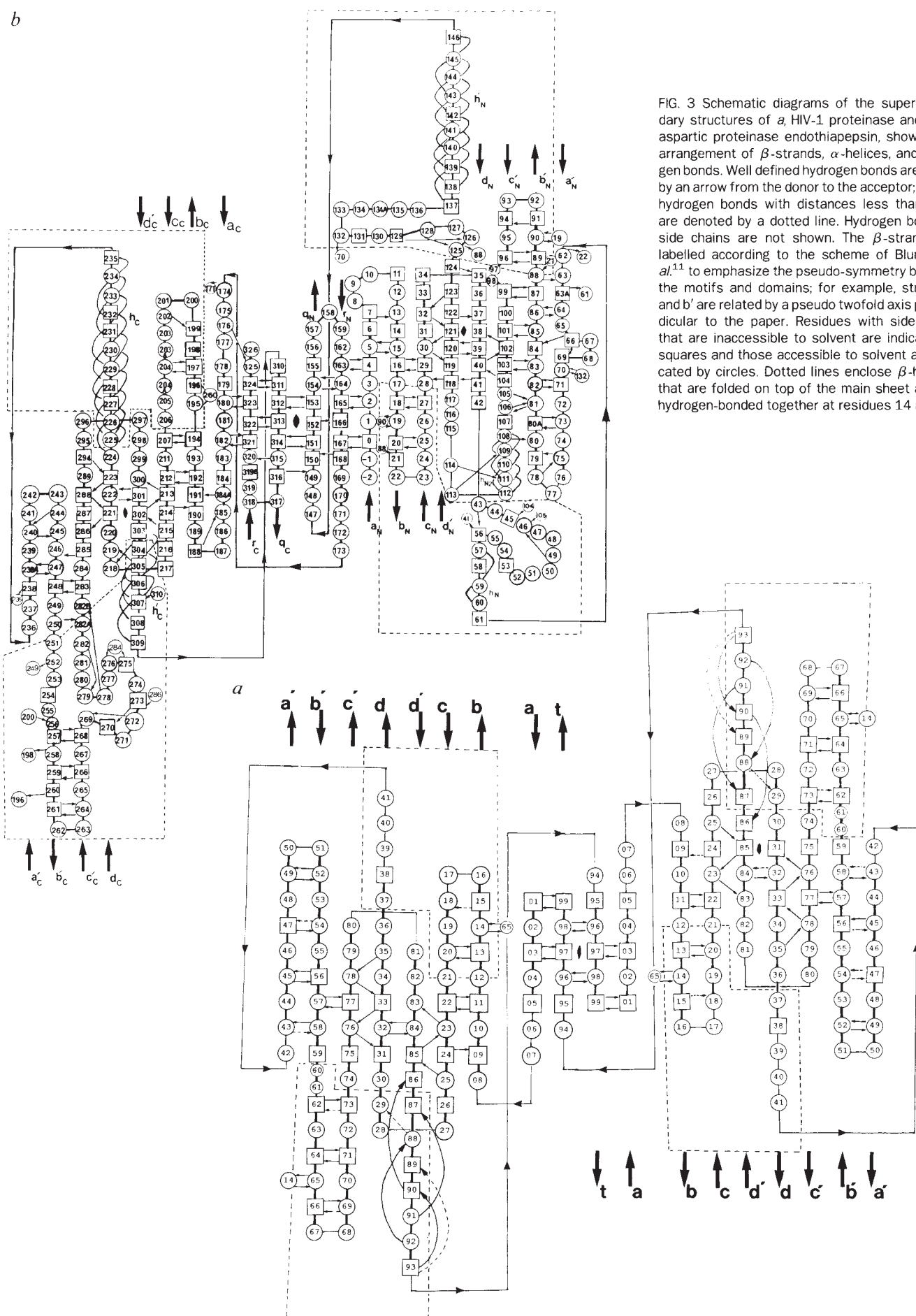


FIG. 3 Schematic diagrams of the super-secondary structures of *a*, HIV-1 proteinase and *b*, the aspartic proteinase endothiapepsin, showing the arrangement of β -strands, α -helices, and hydrogen bonds. Well defined hydrogen bonds are shown by an arrow from the donor to the acceptor; weaker hydrogen bonds with distances less than 3.9 Å are denoted by a dotted line. Hydrogen bonds to side chains are not shown. The β -strands are labelled according to the scheme of Blundell *et al.*¹¹ to emphasize the pseudo-symmetry between the motifs and domains; for example, strands b and b' are related by a pseudo twofold axis perpendicular to the paper. Residues with side chains that are inaccessible to solvent are indicated by squares and those accessible to solvent are indicated by circles. Dotted lines enclose β -hairpins that are folded on top of the main sheet and are hydrogen-bonded together at residues 14 and 65.

Figure 2 shows stereo views of the three-dimensional structure of the enzyme, which exists as a dimer with a well defined and extensive active-site cleft. Figure 3 shows the secondary structure and hydrogen bonding in the dimeric structure compared with that of a pepsin. We adopt the strand nomenclature¹¹ suggested for endothiapepsin. Each subunit comprises two similar motifs related by an approximate twofold axis; each motif contains anti-parallel strands: a, b, c and d for the first and a', b', c' and d' for the second. These are organized together in a distorted sheet, with strands c and d' and strands c' and d forming two pairs of parallel strands (Fig. 3). Strands b and c and strands b' and c' form antiparallel β -hairpins which are folded on top of the first sheet and hydrogen-bonded together around the same twofold axis that relates the motifs in the main sheet. The strands d and d' are followed by an irregular region and an α -helix, respectively, which also lie on the same side of the main mixed β -sheet. Strand a of the first motif is displaced from the main mixed sheet and forms an antiparallel four-stranded β -sheet with the terminal strand, t, and the two equivalent strands from the other subunit of the dimer. These strands occupy the same volume but have different orientations from the six-stranded β -pleated sheet that connects the two domains of the pepsins (Fig. 3b).

The structure corresponds closely to the arrangement of the completely symmetrical dimer that has been proposed as the ancestor of the aspartic proteinases³. The conserved active site residues, Asp 25-Thr 26-Gly 27, form a symmetrical and highly hydrogen-bonded arrangement virtually identical to that described for pepsins^{12,13}. The two threonines are inaccessible to solvent and are hydrogen-bonded so that the O-gamma bonds to main-chain N-H and C=O functions of the other subunit in a fireman's grip. The two aspartates lie roughly coplanar, with their inner carboxylate oxygens hydrogen-bonded to the N-H functions of Gly 27. There is density for a water molecule bound equally to the two carboxylates, as found in pepsins.

An extensive hydrophobic core involving residues indicated by squares in Fig. 3 extends through the dimeric interface. Pro 1, Ile 3 and Leu 5 of the N-terminal strand, and Cys 95, Leu 97 and Phe 99 of the C-terminal strand, form one side of the central four-stranded β -pleated sheet and pack onto Leu 24 and Thr 26 adjacent to the catalytic residues and the hydrophobic residues of the helix, such as Leu 89, Leu 90 and Ile 93. These residues of the core tend to be highly conserved amongst the retroviral proteinases, giving support to this interpretation of the structure. The only truly buried polar residue apart from Thr 26 is Thr 31, which is also hydrogen-bonded to main-chain N-H and C=O functions in a fashion that is often important to protein structure. Thr 31 is either retained as threonine or

occasionally conservatively varied to serine in retroviral proteinases.

One of the significant conservations in the consensus sequences of the retroviral and pepsin-like aspartic proteinases is an amino-acid sequence of two hydrophobic residues followed by a glycine, an example of which is the sequence Ile 84-Ile 85-Gly 86 that is found in HIV-1 proteinase. The retroviral proteinases differ, however, in the presence of an invariant arginine in the next position, for example Arg 87. Such a basic residue is never found in the pepsins. The crystal structure of HIV-1 proteinase shows that this conserved arginine participates in an intersubunit cluster of ionic and hydrogen bonding interactions involving the side chains of the conservatively varied Asp 29 and Arg 8, as in RSV proteinase⁶. No reason for the conservation of these residues is apparent from the structure reported by Navia *et al.*⁷. It is also apparent that the N-terminus is well organized in the dimer and so flexibility cannot be involved in the precursor autocleavage⁷.

The close similarity of the three-dimensional structures of the dimeric retroviral proteinases and the monomeric pepsins suggests a common mode of interaction with substrate. X-ray studies of aspartic proteinase transition-state isostere complexes¹⁴⁻¹⁶ show that the substrate probably binds pseudo-symmetrically in the active-site cleft, using topologically equivalent hydrogen bonding functions and specificity pockets on each side of the scissile bond. By analogy, the carbonyls of Gly 27 in HIV-1 proteinase may form hydrogen bonds to the substrate N-H functions of P₁ and P₂', as suggested by Blundell and Pearl⁸ and by Wlodawer and coworkers¹⁷. The conservatively varied but exposed Leu 23, Ala 28, Val 82 and Ile 84 form the large S₁/S₁' specificity subsites. By analogy with the aspartic proteinases, in which P₃ binds through its main chain to the side-chain oxygen and the main-chain N-H of Thr 219 (pepsin numbering), these interactions would be mediated by Asp 29 in HIV-1 proteinase. This gives further support to the importance of the cluster of charged residues involving Asp 29; the charged residue Arg 87 contributes to the correct orientation of the carboxylate group (Fig. 2b). The pocket, S₃, in this HIV-1 proteinase structure differs from that reported earlier⁷ in that it involves the loop between strands a and b.

The perfectly symmetrical structure of the retroviral proteinases lends credence to the suggestion¹³ that the apparently differing pK_as of the two aspartates in aspartic proteinases are more a function of their proximity than of their differing environments. Nevertheless, the symmetrical dimer must lose its perfect symmetry during catalysis as the substrate has only an approximate dyad. But this is not a restriction for inhibitors, so symmetrical structures may have a major advantage in high affinity binding when compared with an analogous substrate. □

Received 27 July; accepted 3 October 1989.

- Kohl, N. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 4686-4690 (1988).
- Toh, H., Ono, M., Saigo, K. & Miyata, T. *Nature* **315**, 691 (1985).
- Tang, J., James, M. N. G., Hau, I.-N., Jenkins, J. A. & Blundell, T. L. *Nature* **271**, 618-621 (1978).
- Pearl, L. H. & Taylor, W. R. *Nature* **329**, 351-354 (1987).
- Blundell, T. L. *et al. Eur. J. Biochem.* **172**, 513-520 (1988).
- Miller, M., Jaskolski, M., Rao, J. K. M., Leis, J. & Wlodawer, A. *Nature* **337**, 576-579 (1989).
- Navia, M. A. *et al. Nature* **337**, 615-620 (1989).
- Blundell, T. L. & Pearl, L. H. *Nature* **337**, 596-597 (1989).
- Wlodawer, A. *et al. Science* **245**, 616-621 (1989).
- McKeever, B. M. *et al. J. biol. Chem.* **264**, 1919-1921 (1989).
- Blundell, T. L., Jenkins, J. A., Pearl, L. H., Sewell, T. & Pedersen, V. in *Aspartic Proteinases and Their Inhibitors* (ed. Kostka, V.) 151-161 (de Gruyter, Berlin, 1985).
- James, M. N. G. & Sielecki, A. *J. molec. Biol.* **163**, 299-361 (1983).
- Pearl, L. H. & Blundell, T. L. *FEBS Lett.* **174**, 96-101 (1984).

- Bott, R., Subramanian, E. & Davies, D. R. *Biochemistry* **21**, 6956-6962 (1982).
- James, M. N. G., Sielecki, A. R., Salituro, F., Rich, D. H. & Hofmann, T. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6137-6142 (1982).
- Blundell, T. L. *et al. Biochemistry* **26**, 5585-5590 (1987).
- Weber, I. T. *et al. Science* **243**, 928-931 (1989).
- Ratner, L. *et al. Nature* **313**, 277-284 (1985).
- Franke, A. E. *Eur. Patent* 0147178 (1985).
- Haneef, I., Moss, D. S., Stanford, M. J. & Borkakoti, N. *Acta crystallogr.* **A41**, 426-433 (1985).
- Jones, T. A. *J. appl. Crystallogr.* **11**, 268-272 (1978).

ACKNOWLEDGEMENTS. We thank Dr Simon Campbell, Dr Colin Greengrass, Graham Rickett, Mark Ammirati, Duncan Gaskin, Jon Cooper, Dr Jim Pitts and Dr Lynn Sibanda for their help and advice and especially acknowledge the encouragement and valuable comments of Dr Laurence Pearl. This work was supported by the UK AIDS Directed Programme (A.W.), Cray Research (A.H.), the University of Helsinki and the Osk. Huttunen Foundation (R.L.).

The collection connection

Richard A. Fortey

Wonderful Life: The Burgess Shale and the Nature of History. By Stephen Jay Gould. W. W. Norton: 1989. Pp. 346. \$19.95. To be published in Britain by Hutchinson, February 1990.

THE Burgess Shale, which crops out high in the Rockies of western Canada, is a deposit of mid-Cambrian age. It was discovered in the last century by the great American palaeontologist C.D. Walcott, and preserves an astonishing array of soft-bodied animals which allows us a glimpse of the true range of organisms present at an early stage in metazoan history. In *Wonderful Life* Stephen Jay Gould describes the investigation of this famous fossil fauna: "the world's most important fossil animals".

Gould brings alive the monographs and descriptions of these animals which H.B. Whittington and his colleagues have prepared with such care over the past two decades. He presents the unravelling of the morphology of the remarkable finds as an intense psycho-drama, slapping on a bit of hyperbole here, and a colourful analogy there, to buoy the reader through passages that might seem too technical in any other hands. There is no question about the historical importance of the Burgess Shale, and Gould is right when he says that it deserves a place in the public consciousness along with big bangs and black holes.

But the story is directed to a particular end, and it is this end which, to Gould, lends the Burgess fauna its real interest. His contention is that here we see "a disparity in design far exceeding the modern range throughout the world" (p.62). Organisms from the Burgess Shale may belong to phyla unknown in the living fauna; the arthropods — the most impor-

Cambrian, history of life is seen as one of loss of designs — impoverishment, if you like — in spite of the wondrous variety of the modern world. This is surely a dramatic story, but is it true?

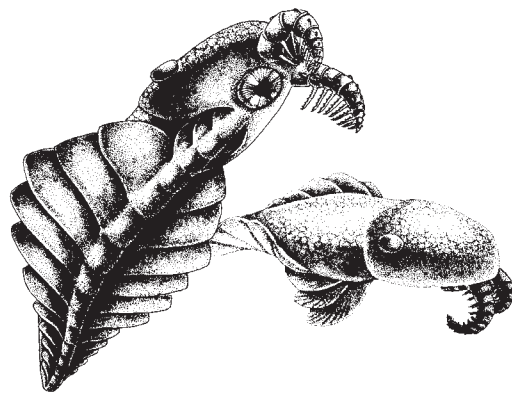
Gould follows a traditional line in taxonomic attitude. The taxonomic status (equivalent to importance) of a group is decided by what he terms 'disparity', which means the morphological 'distance' from other groups. A breakthrough occurred, he maintains, when the Burgess arthropods were recognized as including "a series of unique designs, beyond the range of later groups" (p.168). The most peculiar animals of all might be designated as hitherto unknown phyla, the highest taxonomic division below the Kingdom.

Curiously, Gould hardly mentions the new discipline of cladistics, which was being developed at exactly the same time as Conway Morris, Briggs and Whittington were patiently chipping out Burgess oddities. The cladistic method works by analysing shared similarities, and produces a classification based on derived, homologous characters; and it does not make value judgements about 'disparity'.

The problem with 'disparity' is that its estimation depends on the authority of the expert: how is he to know what makes — what is 'worth' — a phylum? Or what a class? For the Burgess Shale the answer seems to be lack of success in assigning the taxon into a known, living group, so that the fossil becomes 'an animal with no modern relatives'. But unless the animal arose afresh from the mud it *has* to have some modern relatives, be they ever so distant. The end product of the 'disparity' approach seems to be the insouciant recognition of a new phylum, as if it were only to be expected (on p.143 Gould quotes one prominent researcher as remarking, "Oh fuck, another new phylum", on the discovery of a new form).

For systematic affinities to be discovered, the arthropods from the Burgess Shale *can* and should be analysed in terms of the characters they share rather than those they uniquely possess. They seem to fall into a rather ordered array, with *Marrella*, Walcott's original 'lace crab', as the most primitive member, and the trilobites surprisingly advanced.

A similar exercise on echinoderms has seen the positioning of numerous so-called Cambrian 'classes' within a coherent phylogenetic picture. The advantage of this kind of analysis is that it suggests further fruitful questions. For example, the arthropods are metameric organisms, and the characteristic arrangements of appendages (legs, gills and the like) of groups living today may have been assembled piecemeal; the Burgess Shale animals, for all their oddity, may record the stages in this assembly. Hence the fossils may be vital to the understanding of arthropod



Anomalocaris — a "capable swimmer".

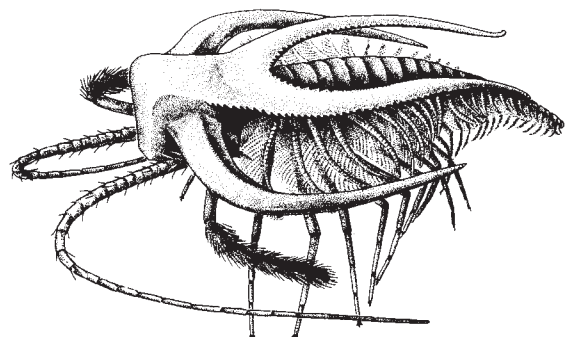
phylogeny; their apparent 'disparity' may partly reflect the process of assembly of advanced clades.

Scientific discussion will centre on the homology of particular characters, which is always controversial in arthropods. But sorting out such complex questions provides more constructive insights into the classification of these arthropods than simply saying: "Gee isn't that weird — it must be a new class!". One might suspect that if, say, the stalked goose barnacles had been known from the Burgess Shale alone ("an arthropod-like animal covered in calcareous plates and with a flexible stalk — goodness gracious!") they would have found themselves in a new phylum. Yet we know from their development that they are specialized crustaceans. It is about time that organisms were classified by characters they share rather than by intuitions about 'disparity'.

Whatever one makes of their relationships, Gould is surely correct in saying that these animals are among the most important for understanding the history of life. Study of them has demanded patience and persistence, and has mostly been carried out with microscope and pencil and paper rather than the high-technology, high-expenditure approach typical of fashionable science. *Wonderful Life* is a compelling story, told with characteristic verve. □

Richard A. Fortey is in the Department of Palaeontology, British Museum (Natural History), Cromwell Road, London SW7 5BD, UK.

Drawings, by Marianne Collins, are taken from the book reviewed here.



Marella — small and elegant.

tant element in the fauna — do not conform to modern groups, with few but important exceptions. Instead we have a time of wild experiment, with morphological components picked like items from a menu in a Chinese restaurant, and placed together in combinations that no longer exist. The subsequent, post-

Dangerous dealings

Alex Comfort

Contamination: A Novel. By Chapman Pincher. Sidgwick & Jackson: 1989. Pp.343. £12.95.

DR Wendy Payne, a microbiologist reluctantly involved in an (unauthorized) British germ-warfare laboratory, whose "effulgence of sex . . . through no conscious effort on her part, could cause stirrings in the private parts of men across the proverbial crowded room", has discovered a bacterial clone which can digest silicon — including that contained in heat-sealed microchips. She is very appropriately murdered on page 1 of this novel after licking an envelope laced with botulinus toxin. The CIA attempt to smuggle a potentially damaging supply of the bacteria in a Star Wars simulator consigned to the Soviet Union: although this is the period of *glasnost* and *détente*, hardfaced GRU plotters, treacherous opponents of the arms race, and contemptible Members belonging to the No Hope Party (the loyal Opposition?) who ask questions about MI5 operations, are all in evidence, and the plot is enacted by the licenced paranoiacs of three nations for 343 pages.

Chapman Pincher's style of writing and his conception of fiction remind me a little of the imitators of the late Nat Gould, but Nat Gould knew a great deal about racing and his backgrounds were plausible. The alarming feature of this extravaganza, which prevents it from being funny, is that Mr Pincher professes to know a lot about the inaptly named Intelligence Community, and his professions might be true. He pictures it as a battery of loose cannon, accountable to nobody, riddled with double and treble agents, capable of any excesses, however hare-brained, and sharing Mr Pincher's rather nutty political opinions. Could he possibly be right?

With the example of *Spycatcher* before them and Mr Pincher's confirmation, scientists might well conclude that intelligence is an ingredient conspicuously lacking from this bunch, and that to hand any interesting finding to such morons under conditions of secrecy is simply not tolerable. I doubt if the author intended it in that way, but this book, for any reader who doesn't give up in exasperation, is one further powerful argument for a Freedom of Information Act. It is easy to disbelieve that there really are people who think and act like this, but the Pincher analysis has confirmation from another 'inside' source several degrees of magnitude up the literary scale, in Le Carré's *The Russia House*.

Even if it were better written, one couldn't recommend this novel as light entertainment to any scientific reader — least of all to any scientific reader who on patriotic or grant-obtaining grounds has inadvisedly committed him or herself to work of 'national importance'. If the Manhattan Project team lived to regret their trust in government, anyone under-

taking similar work today, in peacetime, might well be moved by this insight into Pincherism to write two letters — one of resignation, the other, in a plain unpoisoned envelope, to Mr Tam Dalyell MP, blowing the whistle. He or she would probably be right. □

Alex Comfort is at *The Windmill House, The Hill, Cranbrook, Kent TN17 3AA, UK*.

In the fast lane

John Treherne

Cantor's Dilemma. By Carl Djerassi. Doubleday, New York: 1989. Pp.231. \$18.95. To be published in Britain next Spring by Macdonald.

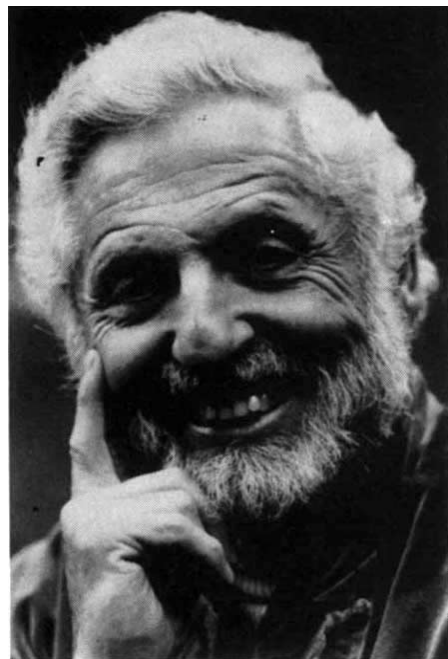
CARL Djerassi can write — there is no doubt about that — which is why reading his novel is such a bruising experience. The narrative bounces along from one repellent character to another. They really are a ghastly bunch.

Take Celestine, for example. At the beginning of the book she seems a relatively innocent little thing. Yet before leaving high school she has been deflowered by the swimming coach and has received her marching orders from the chemistry teacher: "the union card for serious research is a Ph.D. Get it as quickly as possible". To obtain a "top academic position", young Celestine is told — before she even sets foot in a university — that she must infiltrate the old boys' network by doing postdoctoral work "with two different mentors at two different institutions". If things turn out well she will then have *three* men in universities pushing for her. This is all very necessary, for, she is assured, "Chemistry is still a man's world".

Thus counselled, the poor girl takes off for a six-year "combined B.S.-Ph.D. program" at Johns Hopkins. There she works frenetically to further her career, both in the classroom and on her back — it still being a man's world in chemistry — with a dirty old man on the faculty, Professor Graham Lufkin, world expert on insect pheromones and an opera buff. But Celestine knows what she is about ("she felt the quality of intercourse with him — intellectual as well as sexual — was far superior to what her male peers could offer"). They are all like that on the fast track in *Cantor's Dilemma*. Very sportingly for such a cad (he also gives seminars with suggestive titles such as "One-night Stands Among Insects", much to the delight of his prurient audiences), Lufkin eventually sends her to do her PhD research with Jean Ardley. Ardley (formerly "Yardley"; she knocked off the "Y" to get alphabetical priority in multi-author papers) is another fast-tracker who

has had her "tubes tied up" and who, Celestine suspects, was previously bedded by the dreadful Lufkin.

Celestine is next laid by Jeremiah (Jerry) Stafford, of whom she becomes quite fond (despite the fact that he at first lacks "the deft touch of Lufkin"). He is the research student of Professor I.C. Cantor, a big-shot molecular biologist with bushy hair, a large nose and very highly developed ambitions — which include winning the Nobel prize specifically by formulating a



Margo Davis

Djerassi — telling a tale.

new theory of tumorigenesis. As Professor Djerassi puts it: "In the cancer field. . . a generalized theory of tumor production is Mount Everest or K-2. Only superstars climb such peaks, and even they use Sherpas. I. Cantor was such a superstar and Stafford became his Sherpa".

The superstar turns out to be a real dude — for a scientist at least — with his Patek Philippe watch, BMW sedan, silk necktie (worn even in the laboratory), gold Mont Blanc fountain pen, his valuable collection of erotic paintings and expert playing in a Boccherini quintet. Like everyone else in the book he is without a vestige of humour. It is in the quintet that he meets Celestine's Auntie Paula (who is pretty fancy herself — not only does she have expensive antiques and play the cello, but she makes some really helpful remarks about cunnilingus when I.C. shows her his

erotic pictures).

Meantime, Sherpa Stafford is slaving away at hush-hush experiments to prove Cantor's cancer theory. These are so hush-hush and so exhausting that they drive Celestine to complain about his sexual inadequacies, as she soaps his nether regions, using that four-letter word which Ken Tynan — like Carl Djerassi — did not shrink to use. Eventually, Stafford succeeds and all is set for the vital letter to *Nature* which eventually yields the "Nobel in medicine". But not before a lot of worry and unpleasantness about the inability of Kurt Krauss, I.C.'s nasty competitor at Harvard Medical School, to repeat poor Stafford's work. Had Jerry fiddled the results? We are not really sure even when Celestine is disturbed with another boyfriend — appropriately called Roger — to learn that Jerry is to share the Nobel prize with I.C.

Celestine goes along to Stockholm with her Auntie Paula and they all have a wonderful time — pages of it — with both I.C. and Jerry spouting T.S. Eliot in their

speeches and Jerry having a nice gossip with the Swedish Queen about how the Brits use their knives and forks and eat peas.

And that's about it really, apart from some reasonably happy endings — all neatly tied up — and a final bitter letter to I.C. from poor old Krauss, who is still as sick as a parrot about not getting his Nobel prize.

What, I wonder, were Professor Djerassi's motives in writing this book? Is he condoning what he portrays? Is he saying, look, this is how it is. What a smart lot we scientists are: pushy, clever, cutting corners (nudge, nudge), getting a bit on the side (snigger, snigger), swanking about our erotic pictures and Boccherini! If he is, then the hell with it. If, on the other hand, he is saying, look, this rat race is awful, then *Cantor's Dilemma* is a clever, if stomach-turning, way of portraying it. □

John Treherne, novelist and scientist, died on 23 September. This rumbustious review was characteristic of the writer and the man.

Ideal thinking

Steven Rose

Russian Psychology: A Critical History. By David Joravsky. Basil Blackwell: 1989. Pp. 583. £45, \$34.95.

IMAGINE a book entitled *American Psychology: A Critical History*, the first third of which advances that history no further than 1917, the bulk of which follows events only to the early 1930s, and which devotes a mere 30 pages to the years since 1953 — years that have seen such a flowering of the neural and behavioural sciences, and the transformation of our understanding of brain and behaviour. That this disproportion does not seem wholly incongruous for a Western history of Soviet psychology is of course indicative of the extraordinary interest that the first 36 years of the tortured (and torturing) relationship of Soviet communism with science has held for political philosophers, historians and scientists alike. Nor would a history of American psychology of the period (unless of course written by Mr Leo Abse, MP, the recent biographer of Britain's present Prime Minister) be likely to find space for a psychological analysis of F.D. Roosevelt, as this book offers for Stalin. But it also reveals the limits of that interest, for Stalin's death in 1953 is precisely the midway point between October 1917 and the current era, and the second half of this period should surely be as relevant and subject to historical scrutiny as the first for what claims to be, as Joravsky's book does, a comprehensive account.

The superficial view of Soviet science under Stalin sees the period straightforwardly as one in which scientific data and theory were subordinated to political ideology, typified by the Lysenko episode in genetics, often regarded as a simple reciprocal of Nazi eugenics. Such an account misses the passionate concern of Marxism to see itself simultaneously as science and metascience, and the equally powerful determination of an impoverished and backward country to embrace and use science in its desperate push towards modernization. Equally, an unquestioning acceptance of the Western liberal view of science as neutral and above politics, philosophy or ideology misses the opportunity to see, in that Soviet experience, a mirror of some of the problems about the place and organization of science in our own society.

A number of Western scholars have tried to move beyond this simplistic view; the Lysenko affair has been reassessed by Levins and Lewontin, by Lecourt and by Joravsky himself, while Loren Graham's comprehensive *Science and Philosophy in the Soviet Union* has recently been revised and republished. Joravsky's earlier essay in this direction, *Soviet Marxism and Natural Science*, was for a long period after its appearance in 1961 the only book on the theme in English.

The development of Soviet psychology is of particular interest. First, psychology had strong roots in pre-revolutionary Russia, stretching back through Pavlov to Sechenov. Second, the subject matter of psychology has profound bearing on theories of human nature and society, and hence *a fortiori* on Marxism. Third, the theoretical battle between materialism

and dualism, between reductionism and dialectics, which in many areas of science is masked or confined to the philosophers, has been and remains a vital issue of methodology and scientific practice for research in psychology and neurobiology in the West as well as the East. Finally, the vast experiment in social engineering in the Soviet Union during the first half of its post-revolutionary history was directed to the creation of "the new socialist man [*sic*]", and one might well have expected that psychologists would find themselves mobilized to this task.

Thus, from the 1920s on, Soviet neurophysiologists, psychologists and philosophers struggled to develop an adequate theory of the dialectic of brain and society in the production of mind, and an adequate practice of pedology (educational science) and psychotechnics based upon that theory. This struggle with some of the deepest conceptual and methodological problems with which science has to cope, produced responses ranging from the hard-line physiological materialism of Pavlov almost until his last days, through the pioneering work on the social production of mind by the polymath Vygotsky and the later development of neuropsychology by Luria. In the next generation it produced Rubinshtein's activity theory, Anokhin's functional systems approach, the work of Beritashvili, Leontiev and Orbeli, as well as a strong school of functional neurochemistry.

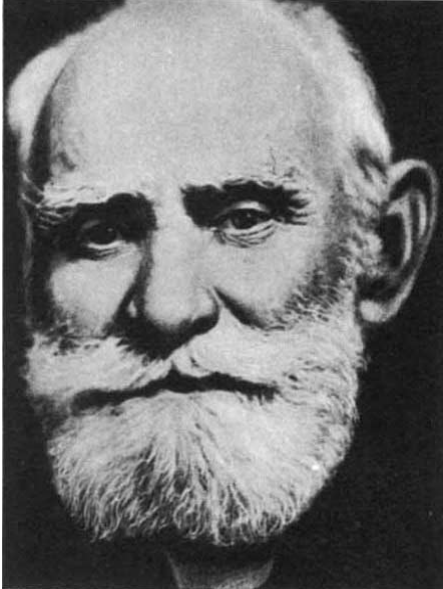
And of course, the work of nearly all of these scientists was punctuated by endless crises, from the desperate economic situation and famine of the early years (to be repeated during the 'Great Patriotic War'), through the increasing tendency to solve scientific disputes by administrative fiat — with all the dread consequences that that entailed — from the 1930s to well beyond the death of Stalin. The debates in psychology seem not to have been quite so literally life-and-death affairs as they were in genetics, although Joravsky lists many victims of the Gulag. True, after 1945 Pavlov was for a time elevated to stardom within Stalin's pantheon, and recantation was forced upon those of his pupils and disciples whose greater understanding of brain physiology and psychology (to say nothing of their better Marxism) led them to deviate from the line of their late master. Even so, the debates in the 'Pavlov sessions' for psychologists, physiologists and psychiatrists in 1950, 1951 and 1952 seem much less staged and more open than the notorious 1948 session of the Agricultural Academy which saw the destruction of genetics and the apotheosis of Lysenko.

Just why this should be so would seem to be a question of some significance — perhaps it was because of the stronger research and philosophical traditions of Soviet psychology and physiology;

perhaps because the declared luminary of the subject, Pavlov, unlike Lysenko, was long dead; perhaps because there was a greater urgency to solve the practical crisis in agriculture than to employ science instead of naked coercion in the production of the new socialist man.

Oddly, however, although Joravsky documents the debates extensively he makes no great effort to address these

Marv Evans Picture Library



Pavlov — developing theories.

broader questions, and the omission is characteristic of a book which in the last analysis yields much less than it promises. Joravsky is not an internalist historian; he is not so much concerned with experimental findings which led Pavlov, Kornilov, Vygotsky or Luria to their theoretical positions, or with the ebb and flow of the scientific debate; rather, he is offering us a metahistory of psychology, an account of the ways in which the leading protagonists of the discipline in the 1920s and 1930s sought to locate their theories within the swirling currents of the party-led attempts to establish a Marxist orthodoxy. In effect this book is not so much a history of Russian psychology but of psychology in Russia.

The author's scholarship is formidable, and the phrenology he has created whilst denied entry into the Soviet Union or access to many key papers is impressive, despite the fact that the book has, on internal evidence, been a very long time in the making and is the weaker for it. Nonetheless, Joravsky creates the strong impression of being in the last analysis rather bored by the scientific questions which dominate the professional lives of the subjects of his history. A reader of this book who opened it knowing nothing even of Pavlov's key experiments and theories would close it being little the wiser, though he would have learned that Pavlov was a vain, opinionated and authoritarian little man whom Joravsky doesn't like very much. Indeed, Joravsky finds it diffi-

cult to observe admirable qualities in almost any of his protagonists. If they don't speculate about the broader significance of their results he finds them narrow technocrats; if they do, they are shallow philosophers, or time-servers who would mould their theories to political exigencies. The treatment — or rather dismissal — of Rubinshtein, Luria and Anokhin in this way is inexcusable in however critical a history. A couple of what Joravsky would probably regard as boringly internalist histories of Russian psychology were published some years ago and do a much more useful job in this regard.

Joravsky would probably respond by saying he wasn't trying to write that sort of book at all. His elegantly worked-over prose, full of ironic asides, musing self-reflections and studied ambiguities may in some ways best be read as an unspoken historiographic debate with what has become known as the Edinburgh school in the sociology, history and philosophy of science, which seeks to reduce debates

about fact and theory in science to a reflection of the interests — the social origins and positions — of the main characters. Joravsky's own interest, as the doyen of the history of Soviet science in the West, remains somewhat unstated. He offers himself as an outsider, neither a sympathizer with nor enemy to communism, an olympian observer and chronicler. But in practice he cannot sustain such detachment; his language constantly lets him down, as when, for all his reading of the Marxist classics, confronted with that robust concept 'working class' he routinely bowdlerizes it to 'lower class'. A real critical history of Soviet psychology, which transcends internalist-externalist boundaries, and helps us to rescue its rational kernel from both Eastern and Western ideological trappings, remains to be written. It will be worth waiting for. □

Steven Rose is in the Brain and Behaviour Research Group, Department of Biology, Open University, Walton Hall, Milton Keynes MK7 6AA, UK.

Pens and petals

Ray Desmond

The Art of Botanical Illustration: The Classic Illustrators and Their Achievements from 1550 to 1900. By Lys de Bray. Christopher Helm, London: 1989. Pp. 192. £25.

THE drawing of flowers — whether for research, decoration or simply for pleasure — is the subject of an impressive literature. A distinguished pioneer in this field was Agnes Arber, whose survey of printed herbals and their woodcut illustrations, *Herbals: Their Origin and Evolution* (1912; 2nd edn, 1986), has become a classic. Another well-known and authoritative text is Wilfrid Blunt's *The Art of Botanical Illustration* (1950; 3rd edn, 1955). His *Illustrated Herbal* (1980), written in conjunction with Sandra Raphael, complements Arber's volume in dealing primarily with the manuscript herbal.

These are just three of many books dealing with aspects of botanical art that have appeared during the course of this century, many of them during the 1970s and 1980s. Artists such as Leonardo de Vinci, Albrecht Dürer, Jacques le Moyne, Georg Ehret, Sydney Parkinson, Pierre-Joseph Redouté, Mrs Delaney and Marianne North have all been the subject of monographs. Other authors have looked at flower drawing and painting from the point of view of the bibliographer and the printer, while exhibitions, museums and auction houses have each generated accounts of specific collections.

Any new book in this area has to be

judged against the high standards set by some of its predecessors. Even though Lys de Bray's *The Art of Botanical Illustration* has no pretensions to original research, aimed as it is at a general readership, and is generously illustrated with colour plates, it is an unhappy piece of work. There are frequent signs of haste in its preparation and the text bristles with errors — misspellings, mistakes in dates of birth and death and in dates of publication — some of which are doubtless due to careless proof reading.

These slips are irritating but others are more misleading. Jacobus van Huysum was employed by Sir Robert (not Horace) Walpole; William Kilburn did not contribute any plates to the *Botanical Magazine*; B. Maund and E.D. Smith were connected with the *Botanic Garden* not the *Botanical Register*; J.N. Fitch lithographed plates for the *Botanical Magazine* but he did not succeed his uncle as artist to the journal. There is no evidence that William Kent landscaped the Prince of Wales's garden at Kew (which was much more than the 11 acres stated by Ms de Bray); Princess Augusta did not extend her property at Kew after the death of Queen Caroline; and W.T. Aiton did not resign from Kew in 1841, he merely relinquished control of the small botanical garden there. This is not an exhaustive list.

It is unfortunate that the author should have chosen *The Art of Botanical Illustration* as her title because it invites comparison with Wilfrid ('Wilfred' in Ms de Bray's book) Blunt's volume. Clearly, though, it poses no threat to the status of the earlier work. □

Ray Desmond is at the Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB, UK.

Adventures of Ben Trovato

Walter Gratzer

Discovering: Inventing and Solving Problems at the Frontiers of Scientific Knowledge. By Robert Root-Bernstein. Harvard University Press: 1989. Pp. 501. \$35. To be published in Britain next month, £27.95.

Fortune or Failure: Missed Opportunities and Chance Discoveries in Science. By Alexander Kohn. Basil Blackwell: 1989. Pp. 199. £14.95, \$24.95.

Serendipity: Accidental Discoveries in Science. By Royston M. Roberts. Wiley: 1989. Pp. 270. Hbk \$19.95, £13.30; pbk \$12.95, £8.65.

Harrap's Book of Scientific Anecdotes. By Adrian Berry. Harrap, London: 1989. Pp. 240. £12.95.

MURPHY'S law, or as the French call it, *le loi d'emmerdement maximum*, dictates that when we drop the breakfast toast on the newspaper or floor it always falls spread-side down. Imagine now recovering the errant morsel and finding, when you turn it over to scrape away the bits of carpet, the long-lost contact lens embedded in the marmalade. This, according to Royston Roberts, is serendipity. Consider next the case of the American professor of statistics, who observed that when his children dropped their bread and butter on the carpet it always landed buttered-side up. A violation of Murphy's law? By no means, for on looking into the matter more closely the professor found that his children were buttering their bread on both sides. This Dr Roberts calls pseudo-serendipity — an unsought discovery that comes to reward the deserving, those whose minds are equipped to receive and interpret the signal, who descry that tide in the affairs of men, which taken at the flood leads on to fortune.

Should you feel impelled to buy these four books, you will learn, not once but four times, that serendipity is a coinage of Hugh Walpole's, derived from the Persian fable of *The Three Princes of Serendip* (that is to say Ceylon, or if you prefer, Sri Lanka), meaning the happy faculty of making unexpected and fortunate discoveries. Robert Root-Bernstein explores the phenomenon in the greatest depth and his is by far the most ambitious work. It is cast in the form of a running seminar, for six voices, about the nature of the scientific process. One of these, referred to as Imp, I take to be the author's *alter ego*. The discussion ranges over the practice, philosophy, organization and sociology of science with no little erudition, even if one occasionally fancies that one is being beaten about the cranium with it ("As Paul Valéry says . . ."). Imp and his friends develop many a provocative but abundantly documented thesis. Science, they find, for example, is getting worse: ever more scientists are generating less and less new knowledge per head; the success of science at the national level does not reflect the munificence of its funding. Many small countries — Den-

mark is one — do better than the United States, let alone the Soviet Union, by such criteria as Nobel prizes per head of population.

Root-Bernstein's protagonists argue a great deal about creativity. How may it be promoted and can it be taught? The uneasy outcome seems to be that it can be inculcated by educational precept,



Contemporary mistrust of Jenner's theory.

according to such formulae as used to be advertised by Edward de Bono; the existence of the great scientific dynasties — Nobel laureates begetting Nobel laureates — suggests, on the other hand, that it is better learned by example, as part of the patrimony of a great laboratory.

Root-Bernstein illustrates his argument with many examples. He favours research by bold and hazardous flights of the intellect: to be audaciously wrong is better than to be timidly correct. ("This theory isn't right", Debye was supposed once to have declared after a seminar, "it isn't even wrong".) Imp is ever keen to challenge the wisdom of the day. Crick's Central Dogma is his favourite target. But here the urbane mask suddenly slips, revealing, as it seems, a bloodshot glint of messianic fire; for it is Root-Bernstein's own work (published in the *Journal of Theoretical Biology*) that his surrogate cites in evidence. He is out to black the Establishment's eye by breaching the infamous Central Dogma with a demonstration that information can flow from protein to protein. Model building leads him to a scheme of complementary amino acids. In search apparently of a spot of serendipity, he meanders down a byway

and arrives at a model for an interaction between a neurotransmitter, serotonin, and myelin basic protein. Imp speaks highly of Root-Bernstein's work. He then goes on to denounce the intolerance of the guardians of entrenched opinion towards such enterprising conceits as four-stranded DNA structures, and the purblind reluctance of grant committees to shell out money for the pursuit of heterodoxy. I find this slightly queasy reading: it seems to me to come close to the cry of the crackpot down the ages, demanding a blank cheque to support what Irving Langmuir called (with criteria by which it could be recognized) pathological science.

Root-Bernstein quotes with approval Albert Szent-Györgyi's remark about the fatuity of a system of funding that required of him some indication of what he proposed to do with the money. If, he maintained, he knew what he was going to discover, there would be no need for him to discover it. Judged by this argument of course all discoveries are by their nature serendipitous. But Root-Bernstein goes on to make a compelling case that the grant system is becoming increasingly rigid, and acts to discourage those digressions in pursuit of the unexpected that have so often led to the most striking discoveries. His analysis of conceptual upheavals in science over the past century or so also leads him to the conclusion that a large proportion have come from unknown researchers, operating alone in the geographical wilderness. If he is right then the cultivation of 'centres of excellence' and the projected elimination of research from minor seats of learning in Britain are a dangerous folly. Our scientific mandarins should examine his evidence with care.

Another of the contentions ventilated by Imp and his companions is that the creative process is nourished by artistic imagination, and that the scientific elect most often possess accomplishments in music, in painting or in poetry. The evidence is marshalled in a table that spreads itself over several pages. This may be right, but for my part I have too often found the giants of our profession to be boorish thugs to be wholly convinced. Moreover, while there may be no doubt about Einstein's love of music, his proficiency on the fiddle by all accounts did not far outstrip that of Heifetz in cosmology. (His frequent partner, Martinu, who wrote a sonatina for him, was supposed to have snapped in exasperation: "Damn it, Einstein, can't you count?")

But enough grousing, for Root-Bernstein has written an absorbing, stimulating, sometimes infuriating book, full of entertaining ideas, and everyone with an interest in science and its ways will benefit from reading it. ►

Alexander Kohn and Royston Roberts plough the selfsame furrow. All the old tales are here, nor do they fade with the telling, and there are also a good many I did not know. Both miss the story of Dr Murphy's patient in Boston, who was fond of raw liver and became cured of pernicious anaemia, but not much else that I could think of. Kohn concentrates more on medical discovery, and in today's world of mandatory LD50 tests, clinical trials and FDA approval, the casual manner in which dire procedures were tried on human subjects — the family, often enough, of the experimenter — taxes credulity. Jenner, it may be remembered, "with great fortitude tried the cow-pox on his wife and children", but before that he had taken the precaution of inoculating a farm-boy, James Phipps, and then injecting him with the real article — a smallpox extract. Much more recently a research director at Boehringer, screening a series of compounds for vasoconstrictor activity to use against the common cold, tried one out on his snuffling secretary; she promptly sank into a sleep that lasted 24 hours, and her blood-pressure was found to be alarmingly low. By this means clonidine, the sovereign treatment for hypertension, was discovered.

In the drug field, especially, the exercise of reason seems to have played little part in the best discoveries. Lithium therapy is an example to make any rational pharmacologist put by his smoked drum and seek a tranquil retirement. Kohn tells the story well: an Australian psychiatrist conceived the notion that manic disease was caused by a toxin, and that this would appear in the urine. So he injected the urine of his patients into guinea pigs. The guinea pigs reacted badly, but less so, he judged, to control urine. The next stop was to identify the toxic principle. Urea killed the guinea pigs, but its concentration was the same in manic and normal urine. How about uric acid then? But free uric acid is insoluble, so the dauntless experimenter prepared the lithium salt. Not only did it appear innocuous, but it counteracted the toxicity of urea and exerted a 'calming effect' on the agitated rodents. To do the psychiatrist justice, he did then ask himself whether it was the uric acid or the lithium that was exerting the benign affect, and indeed lithium carbonate proved equally soothing. He now tried lithium carbonate on a patient, with miraculous results. A paper was written and ignored until it was discovered in the fringe literature by a Danish worker some years later. Lithium therapy became in due time one of the

In Memoriam

For "quaker", pray, and for "moth-eaten", mourn.
This threnody recalls their fate forlorn.

We catered to their reproductive need.
We cherished and encouraged them to breed.
We nourished them (for 15 cents a day)
Until, on orders, they were shipped away:
Their natures chromosomally imbued
With genes for "non-immunity" or "nude",
From whites to pure C-57 blacks:
A thousand, thousand cultivated JAX.

In two score years and two, that realm of mice,
Felled once by fire, has now been stricken twice,
Ignited by explosion of propane
To holocaust in far Bar Harbor, Maine,
Where, cage on cage, in box on burning box,
There perished scores of rare foundation stocks.
On shelf and bench, in closely serried racks,
There died a pyre of martyred murine JAX.

Weep, then, for mice in Roscoe Jackson's name,
Incinerated in a burst of flame.
Like fair Brunhild', by fiery circle girt,
They died, the noblest breeds of Mount Desert.
Their bodies ashed, their genes distilled away —

Mourn for "moth-eaten", and for "quaker", pray!

Ralph A. Lewin

great triumphs of clinical science; but manic patients have no toxin in their urine and the guinea pigs became quiet only because the lithium made them sick.

Roberts is an organic chemist and his trade is perhaps the richest of all in fortuitous and perverse discoveries. He has written an entertaining book, accessible to all. Here is Dr Sapper, who breaks a thermometer in the reaction vessel and is rewarded by the discovery that mercuric ion is a catalyst for the oxidation of naphthalene to phthalic anhydride, and so begets the modern dyestuff industry. Roy Plunket, in search of a better refrigerant, sets out to try a prep with tetrafluoroethylene. The weight of the gas cylinder shows it to be full, but nothing comes out. On being sawn open it yields up a waxy solid, now known as Teflon. A negligent graduate student leaves his cigarette smouldering on the bench while he works; he takes a drag and finds that it tastes intensely sweet. He has discovered saccharine. Forty years later James Schlatter at Searle, pursuing ulcer drugs, is synthesizing a terminal dipeptide of gastrin. He licks his fingers and experiences a sensation of sweetness. He tastes his dipeptide ("It was a bold man who first swallowed an oyster", said King James), and lo, aspartame or Nutrasweet is the fat man's friend. And so it goes.

Who are these holy innocents, whose curls fortune ruffles with her fingers, while giving the rest of us the flat of the hand

across the windpipe with the other arm? Roberts believes they are just like us, only quicker of wit and sharper of perception. For every one of them there must be a thousand others, who failed to take the tide at the flood, and the voyage of whose lives was accordingly bound in shallows and in miseries.

Kohn enlarges on the subject of missed opportunities. These tales are inevitably harder to come by, but he tells for example of a well-known biochemist, still among us, who went to the wrong seminar and failed to initiate a collaboration which would probably have started recombinant DNA a year or two early. He might have mentioned also the Japanese physicist, who shelved an observation that should have led him to the discovery of the neutron; it was said that in his later years cruel students would drop the word in his presence to see the tears well up and course down his cheeks. But what of all the discoveries and inventions that have not been made? Like children unconceived, they will never be missed, but one can imagine them massing in the shadows to mock the bureaucrats who work so hard at

keeping scientists away from the paths of dalliance.

On his deathbed, Lord Camrose was supposed to have addressed the following exhortation to his assembled family: "If you want to die as rich as I am, never brighten up *The Daily Telegraph*". But do not let this put you off *Harrap's Book of Scientific Anecdotes*, which comes from the pen of the *Telegraph's* science correspondent. The book is not what its title suggests, but rather a collection of Adrian Berry's favourite passages of scientific and technological journalism, yarns of discovery and exploration from *The New Atlantis* to Chernobyl. It is agreeably written, spiced with witticisms and altogether good fun.

I shall leave you with just one example of Berry's historical research. It concerns the reception given to Alexander Graham Bell's telephone. The mayor of an American city grasped its potential at once: "One day", he said, "there'll be one in every city". But what did the experts think? Here is the Chief Engineer of the British Post Office, reporting with considered wisdom to a House of Commons committee: "Americans have need of the telephone — but we do not. We have plenty of messenger boys". So much for peer review. □

Walter Gratzer is in the MRC Cell Biophysics Unit, King's College London, 26-29 Drury Lane, London WC2B 5RL, UK.

Sunk without a trace

June Goodfield

The Mind Has No Sex? Women in the Origins of Modern Science. By Londa Schiebinger. *Harvard University Press*: 1989. Pp.355. \$29.50, £23.50.

"LET US now praise famous men, and our fathers that begat us", runs the verse from Ecclesiasticus. Not, you notice, that women or mothers were in any way involved nor were they likely to have been present in the writer's mind even as he concluded: "And some there be which have no memorial. Who are buried as though they had never been". It was ever thus, I concluded, as I finished Londa Schiebinger's remarkable book.

Its title comes from an assertion made over 300 years ago by Francois Pollain de la Barre, as part of an attempt to secure a place for women in the then emerging modern science. The author of this book adds a question mark to his phrase — not because she has any doubts about the matter but, I believe, because over the centuries thousands have stated equally assertively that the fairer sex is also the weaker, both mentally as well as physically. And they have appealed to a whole range of rationales — themselves socially and culturally determined, subject equally to the whims of fashion or assumption — to underpin their preferred judgements. By now no one, I trust, denies that during history women have been largely excluded from scientific institutions (as indeed they have from almost all other institutions) and the mechanisms of exclusion, whether structural, sexual or derived from supposed gender differences, are still largely in operation. Schiebinger's intention is not to add to the burgeoning literature on this topic but to try to understand what these gender differences are, and how they operate in science today. Her chosen method is to examine the past — the revolution in European science in the seventeenth and eighteenth centuries to be precise — in order to answer this problem: what role did the "women question" and the debate over "female nature" play in the origins of modern science?

This is a formidable task, but it has been undertaken by someone of formidable scholarship who is blessed with one additional gift: she understands that social and historical dynamics are highly complex yet has remained undaunted. Halfdan Mahler, when Director General of the World Health Organization, commented recently that "development is a very messy affair" and added "God protect me from the simplifiers". The point applies to this issue too. If we are to make any progress at all in

understanding the present inequities, let alone rectify them, we need to eschew simplicity and generalizations and do full justice to the underlying human complexity.

Accordingly, the author approaches her analysis from four different angles but always with the question of gender to the forefront — that is, are men and women different by nature or by nurture; if so, what difference does (and did) this difference make? She explores in turn, first, the institutions of science which in the period under discussion were fostered in a variety



Mary Evans Picture Library

Labour pains — midwives became auxiliaries in a male-dominated medical system.

of settings (academies, Parisian salons, princely courts and craftsmen's workshops); second, the lives of some women, predominantly from the aristocracy and the crafts, who pressed against the boundaries of convention to practise science; third, how the biological sciences traditionally merged sex (a matter of genes) with gender (a matter of culture, education and expectations) to produce supposedly scientific readings about the female nature that were used equally by those who argued for, or against, women's full participation in science. Finally, her gaze is directed towards the cultural meanings of femininity and masculinity, and how a woman's ability to do science has seemingly become irreparably meshed into issues of gender.

No one of these elements is, she rightly insists, a causal agent that alone determines where women are in science today. Each, however, had a part in the process, and in the dynamic that determined the present outcome they have all played important and interdependent roles.

The extent of research and the material uncovered is remarkable. Where and how did Schiebinger find it all? For this is intellectual archaeology at its very best, and a landscape is laid out for us the existence of which most readers will be totally igno-

rant. There are some surprises along the way: matters were not totally ever thus.

Although the extent of women's activity, success and penetration into science varied from class to class and country to country, their ultimate exclusion was not a foregone conclusion. Aristocratic women dominated the discussions in the salons of seventeenth-century Paris; women artisans did notable astronomical and agricultural research in Germany; the Enlightenment was a great time to be alive, when women's desire to be involved in science was assisted by many social trends and even the bastions of some of the new scientific academies were temporarily breached. Yet time after time all such efforts were stymied and it is ironic that the ultimate *coup de grâce* was to come from the very enterprise that women tried to develop. For modern science, too, finally came to codify a view of women that set the seal on their exclusion.

But there is, of course, more than one way to skin a cat, as the history of midwives illustrates. Since ancient times this 'craft' group had peacefully co-existed alongside medical men, each with their own art and monopoly. But both the peace and monopoly were to be shattered in the two centuries in question. Whereas doctors, surgeons, apothecaries, dentists and veterinarians gradually transformed themselves into self-governing, self-regulating professions, midwives never succeeded in doing this though they certainly tried. But still they remained regulated by men — by the King's chief barber-surgeon in Paris, by the clerics in England. Finally, the *accoucheur* appeared — the man-midwife who was now called in for routine cases, not just abnormal labours. By the end of the eighteenth century, midwives were just auxiliaries to a male-dominated medical system: the practice of obstetrics and gynaecology passed over to men.

This example is telling for it illustrates the nature of one of the Catch-22 traps that ensnared so many women who wished to progress in science. Midwives could not learn new methods or practices because they could neither attend universities nor establish their own medical colleges; they could not do those things simply because they were women. No matter what arguments they employed, what authorities they quoted — from the Bible to the new ethnology — to try to secure their ancient privileges, it was hopeless. And when finally they were driven to the argument of 'natural propriety' matters were made even worse. For 'natural innateness' was the final appeal that, above all, kept the women at the hearth in their traditional child-bearing roles.

In giving this account of the role of science in the creation of gender, backed by this mass of fascinating detail, intriguingly interpreted, Schiebinger has truly

travelled uncharted territory. But has she really answered her own question? Societies "intent on preserving the sharp social and intellectual distinctions between the sexes" have, for the most part, deliberately cultivated perceived sexual differences. But what are these? Do they indeed exist; if so, what difference does the difference make?

Here these questions are posed rather than answered. Unravelling the multifarious strands of women's role in the history of modern science is an essential first step, brilliantly executed in this book. It is to be hoped that this study will pro-

voke further analysis into the contemporary scene and present powerful trends in science. For every Maxine Singer, now President of the Carnegie Institution of Washington, there are still hundreds of others striving to make their mark in the scientific enterprise in conditions where Catch-22 traps still operate. Will it be ever thus? Sometimes I'm optimistic, but reading what happened in history, and not yet discerning overpowering new trends, sometimes I wonder whether such optimism is justified. □

June Goodfield is at The Manor House, Alfriston, Polegate, Sussex BN26 5SY, UK.

Asking questions

Igor Aleksander

Ideas and Information: Managing in a High-tech World. By Arno Penzias. W.W. Norton: 1989. Pp. 224. \$17.95, £12.95.

FOR many people, phrases such as 'the information age' or 'high-tech world' evoke feelings of the hackneyed, the over-sold. Not so for Arno Penzias, who sees 'managing in a high-tech world' as a matter of exciting discovery and wonderment. In *Ideas and Information*, Penzias, a Nobel laureate for physics in 1979, glances over his own specialist fence at the nature of information technology and the effects it has had on him and the world he inhabits. The world 'managing' in the subtitle is used in the sense of 'coping': Penzias is a technology enthusiast. His book is intended to give heart to those who feel that coping in a world of faxes, electronic mail, word processing, databases and the like is getting beyond them.

The style is anecdotal without being trivial, authoritative without being overpowering. And it *has* to be enthusiastic because, after all, Penzias is the vice-president for research of one of the world's most renowned information-technology companies: AT&T's Bell Telephone Laboratories. The anecdotes add fun and a personal touch to what could so easily otherwise have been yet another dull catalogue of the benefits bestowed on mankind by the advent of computers. Penzias also goes to great lengths to distinguish the rule-enforced computer from the free-learning, exploration-loving human being. I suppose that there are still those who get the two confused to the extent that such descriptions are necessary.

The book is written with an implied sense of history: the history of symbols, the history of mathematics and the history of computing machines. Penzias argues that the manipulation of symbols distinguishes computers from other machines in the way that being able to use symbols distinguishes human beings from animals.

Symbols were valued by the prehistoric painters of the images in the Lascaux caves in France, by the Sumerians of 3300 BC who invented information storage (tokens in a clay jar), and by the Egyptians and the Chinese, who raised application of their intelligence to higher planes by inventing numbers and mathematics. So, for Penzias, the computer is the ultimate celebration of human intelligence: a machine that embodies the finest points of the use of symbols and manipulations.



Arno Penzias — "cheerful optimism".

He does not confuse person and machine, however, as seen in this almost thrown-away footnote: "As we explore the computer's ability to match human information-processing capabilities, we shouldn't overlook the fact that these machines have much to offer society without emulating humans". At the same time as he creates the distinction between living and mechanical information processors, he pinpoints the very factor which, if mechanized, could reduce this crucial distance: "On the other hand, I am sure that this situation will change drastically when (and if) we can devise a computing architecture that learns from examples, the way humans do, rather than merely obeying instructions. Nature produces countless billions of such computers in the brains of living creatures". In these sentences he has identified (perhaps unwittingly) the scent that the contem-

porary hounds of neural-network research have sniffed out, and that is leading them towards learning computers with which they will try to revolutionize computing by narrowing the gap between the mechanical and the biological.

The mix of anecdote, history and insightful comment is a characteristic of the book as a whole. Penzias describes milestones in the making of information machines such as Chomsky's analysis of language, Hopfield's revival of neural nets and Joseph Weizenbaum's totally mechanical 'psychiatrist emulator', Eliza, which makes its interlocutors think that it 'understands' their problems, while it is simply throwing their own words, modified by a few simple rules, back at them.

For me, a high point is reached in Penzias's recounting of I.I. Rabi's description of the reasons for his (Rabi's) success, also crowned by the Nobel prize in physics. "Have you asked any good questions today?", his mother would ask him on his return from school. This, surely, rather than just the ability to learn, must be the factor that best distinguishes organisms that possess intelligence from those that do not. The ones that do ask very good questions.

There are some unfortunate blemishes. Bell Labs is presented as the source of all discovery, which will irritate readers who happen to know that not all high-tech inventions were made at that institution. For example, the comment "At Bell Labs . . . a robotics research group has built a robot that can follow spoken instructions" is followed by a description of work that was done by Terry Winograd at the Massachusetts Institute of Technology in the early 1970s. There is no mention of Winograd. Even Hopfield is described as a Bell Labs' physicist, giving no credit to the fact that he is also an academic at Caltech.

But perhaps the worst sin is that there is no mention of the down side of computers: the overrating of technology for its own sake; the fact that the purchase of information technology in a badly managed organization can amplify the bad management; the need to change one's behaviour to meet the needs of computer-driven procedures (as in banks); new opportunities for fraud; the possibility that people may be forced to become less skilled as well as more skilled. For Penzias these difficulties do not exist, or if they do he may have decided to ignore them for fear of damaging the up-beat tempo of his book. Without a doubt those who need reassurance about coping in the information age will be rewarded by the cheerful optimism purveyed in *Ideas and Information*. Why spoil it with bad news? □

Igor Aleksander is in the Department of Electrical Engineering, Imperial College, London SW7 5BT, UK.

After the two cultures

Adam Kuper

Academic Tribes and Territories: Intellectual Enquiry and the Cultures of Disciplines. By Tony Becher. *Open University Press*: 1989. Pp.200. Hbk £27.50, pbk £9.95.

TONY Becher's enquiry into the mores of academics is a belated reaction to C.P. Snow's division of intellectual life into two great camps, arts and science, which (if they ever existed) have now long been abandoned for smaller, more intimate, redoubts. An anthropologist, Clifford Geertz, has pointed out that "most academic communities are not that much larger than most peasant villages and just about as ingrown", with their characteristic totems, initiation ceremonies and ancestor rites, modes of work, codes and systems of authority. Professor Becher pursues Geertz's suggestion that these societies should be treated rather in the way that anthropologists study exotic tribes, as social units that have distinctive cultures.

For his enquiry he selected 12 disciplines — biology, chemistry, physics and mathematics, representing the pure sciences; mechanical engineering and pharmacy, as applied sciences; economics and sociology, from the social sciences; history and modern languages from the humanities; and law, representing the social and administrative professions. It is immediately obvious that these categories are not watertight (mathematics, economics and history could all be reclassified), and Becher's thesis is in fact that the most significant criteria for classifying academic pursuits cut across not only the science-arts divide but also disciplinary boundaries, and that they define most clearly the subdisciplines which we call specialisms. This is where the academic has his true being.

Becher's main criteria are familiar: hard versus soft, and pure versus applied pursuits. Hard specialisms use quantitative rather than qualitative methods, clearly define problems and deliver cumulative additions to a body of knowledge. They also tend to be free of value conflicts. Soft specialisms use qualitative methods, address diffuse issues which are often dictated by social concerns (and therefore invite political controversy) and they do not usually yield a body of uncontested knowledge. A more original distinction is that between 'hard, applied' specialisms, which proceed by trial and error, and deliver products which are assessed in terms of their efficiency, and 'soft, applied' professions which use case law,

are evaluated in political and moral terms, and produce protocols rather than instruments. Most interesting is Becher's distinction between 'rural' and 'urban' specialisms, the urban specialism being one in which large numbers of workers, operating in competitive teams, address the same closely defined problems, publishing rapidly and often informally, while the rural worker practises his individual craft in a village workshop, where the postman seldom calls.

By applying these criteria to subdisciplines, Becher captures something of the perceived reality of academic life, where the most sensitive boundaries are often drawn within the discipline. He quotes biologists on biologists — "microbiologists are narrow, but geneticists are lively"; "zoologists are more adventurous and less conventional than botanists" but tend to be "arrogant and overconfident"; botanists "hide behind herbarium cases and hate one another". These stereotypes point to some real differences between specialisms which have to do at once with the type of work involved and with the type of people who choose that sort of work (or who are moulded by what they do).

The cultural peculiarities of the tribe are also a matter of organization. Careers, for example, follow characteristic trajectories in particular disciplines. The phenomenon of burn-out, the scientific menopause, which pushes middle-aged men into administration, or philosophy, is most characteristic of hard, pure specialisms. In soft, pure fields scholars become more expert with time, so that a historian may expect to reach his peak only in his fifties. Publications take a different form in different fields. Soft, pure fields typically require books, while hard, pure fields know only short papers. The former are typically written by one scholar over many years, the latter by teams of collaborators on short-term projects. Soft disciplines tend to be dominated by a small number of Grand Old Men, the hard disciplines are multi-centred, throwing up charismatic leaders with shorter tenure and more rivals.

Of course, all academic disciplines also share in the general academic culture, about which Becher has some useful reminders, starting with one of the universals of academic life, "that nearly everything is graded in more or less subtle ways".

His central premise is that "the main currency for the academic is not power, as it is for the politician, or wealth, as it is for the businessman, but reputation". Academic reputation depends on the quality of publications, but it is also affected by institutional affiliation. Publications emanating from famous laboratories or think-tanks start life with a silver spoon in their mouths. Reputation in turn brings

Selected titles from Kluwer

Climate and Geo-Sciences A Challenge for Science and Society in the 21st Century

Proceedings of the NATO Advanced Research Workshop held in Louvain-la-Neuve, Belgium, May 22-27, 1988

edited by **A. Berger, S. Schneider and J. Cl. Duplessy**

NATO ADVANCED SCIENCE INSTITUTES SERIES
C: Mathematical and Physical Sciences 285

This book presents a series of expert reviews of recent progress in many climate-related disciplines. The future is also signposted, with contributions on emerging technologies.

1989, 728 pp.
Hardbound Dfl. 265.00/£88.00/US\$134.00
ISBN 0-7923-0404-7
Paperback Dfl. 140.00/£46.00/US\$69.00
ISBN 0-7923-0412-8

Artificial Intelligence: Its Scope and Limits

by **James H. Fetzer**

This work explores the philosophical foundations of AI from the perspective of the theory of knowledge. It makes a sustained, systematic and original contribution to defining the nature of this field.

1990, 325 pp.
Hardbound Dfl. 160.00/£53.00/US\$69.00
ISBN 0-7923-0505-1
Paperback Dfl. 69.00/£23.00/US\$29.00
ISBN 0-7923-0548-5

Analog VLSI Implementation of Neural Systems

edited by **Carver Mead and Mohammed Ismail**

THE KLUWER INTERNATIONAL SERIES IN ENGINEERING
AND COMPUTER SCIENCE 80

In *Analog VLSI Implementation of Neural Systems* Carver Mead and Mohammed Ismail present leading research contributions encompassing the entire range of topics currently under study in this exciting new discipline. All contributors' descriptions include detailed characterization of a working chip. Furthermore, no design has previously been published. As a result, this volume presents the most recent innovative work in this emergent field. Each contribution devotes considerable attention to design shortcomings, as well as to notable successes achieved.

1989, 272 pp. ISBN 0-7923-9040-7
Hardbound Dfl. 150.00/£39.25/US\$55.00

Transition Metal Complexes as Drugs and Chemotherapeutic Agents

by **Nicholas Farrell**

CATALYSIS BY METAL COMPLEXES 11

Aspects of both chemotherapy and pharmacodynamics are included and the book unites for the first time in one volume a description of the wide range and applicability of metal complexes.

1989, 304 pp. ISBN 90-277-2828-3
Hardbound Dfl. 180.00/£59.00/US\$89.00



**KLUWER
ACADEMIC
PUBLISHERS**

P.O. Box 322, 3300 AH Dordrecht,
The Netherlands

P.O. Box 358, Accord Station, Hingham,
MA 02018-0358, U.S.A.

Reader Service No.38

power over the careers and reputations of others. Within the political systems of disciplines and specialisms, peer groups may serve as democratic controls, particularly in reviewing, but power tends to concentrate in very small networks.

Suggestive though it is, this short study is nevertheless only a preliminary excursion into the field. Becher restricts himself to élite institutions. His fieldwork was done in Britain and the United States, but he gives an undifferentiated account of the Anglo-American university world, although the national variations are surely very striking. A future study, more sensitive to the national character of academic enterprises, should also consider Dutch, German or Scandinavian universities, with their civil service cultures, or European think-tanks, like the Max-Planck Institutes, on which *Nature* has been publishing some unsettling ethnographic reports. A reading of Pierre Bourdieu's *Homo Academicus*, a dense and original recent study of the French academic system, will convince the foreigner that they do things very differently in France. For the purposes of Becher's argument, it is, for instance, noteworthy that the prestige hierarchy of disciplines diverges from the Anglo-Saxon model, and also that the leading research teams are to be found outside the university sector.

Becher does not draw together his scattered comments on the connections which might be made between the peculiarities of various academic tribes and the nature of the territories they inhabit. Here again a comparative approach would have suggested fresh questions. Are German physicists more like German economists in their career patterns and organizational hierarchies than they are like American physicists? Or is it something about physics, or economics, that is decisive? Professor Becher briefly discusses the different environments of the various disciplines — the market pressures on engineers, the professional demands made of lawyers and so forth. But these also vary between different national political systems and markets, as may be seen from the contrasting views of academic engineering in Britain and in Germany.

One important implication of the study is nevertheless clear. These various enterprises are not readily measured by the same criteria because they differ so markedly (and probably necessarily) in their goals, methods, work habits and relationships to the outside world. Bureaucrats will have to refine their measures of efficiency if they are to do justice to this diversity of academic society. □

Adam Kuper is in the Department of Human Sciences, Brunel University, Uxbridge, Middlesex UB8 3PH, UK.

Gossip value

Lewis Wolpert

Life Among the Scientists: An Anthropological Study of an Australian Scientific Community. By Max Charlesworth, Lyndsay Farrall, Terry Stokes and David Turnbull. *Oxford University Press, Australia: 1989. Pp.304. Pbk A\$19.95. To be published in Britain and the United States next year.*

A MEASURE of one's understanding of other groups is how well you appreciate their humour. This applies as much to scientists as any subset of people, for jokes and wit are a prominent feature of laboratory and institutional life. Yet humour is



Macfarlane Burnet — Nobel winner.

not mentioned in this book. The scientists at the Walter and Eliza Hall Institute of Medical Sciences in Melbourne — the group considered — may of course be unusual, but a more likely explanation is that the social scientists who carried out this anthropological survey simply failed to understand much of what they were studying.

The authors, who have adopted the curious convention of writing in the first person, like a conflated committee, set out with the provisional hypothesis that scientific knowledge is socially 'constructed'. This is a fashionable view among many sociologists of science and implies that science has rather little to do with the real world that scientists study, but instead is shaped by a complex set of social factors. This relativist stance is then used to argue that scientific knowledge has no particular virtue or validity. These authors take a more equivocal position but basically stick to this idea, and the possibility that science provides understanding of natural phenomena is virtually ignored.

The Walter and Eliza Hall Institute is very distinguished, its contributions to science covering a wide range of topics in molecular and cellular immunology. The first director (1944–1965), Sir Macfarlane Burnet, won a Nobel prize for his clonal selection theory and in the book there is an attempt to view him as a local hero whose social function is to legitimize a particular conception of science — the idea of an individual scientist making a discovery which revolutionizes the field. Burnet was hostile to molecular biology and his general style is contrasted with that of the current director, Sir Gus Nossal. But the analysis provides no insight into how Burnet came to make his great discovery or into the basis of his antipathy to molecular biology. It may not have been his dislike of complex apparatus and expensive equipment, as Charlesworth and his colleagues suggest, but rather his inability to see the power of molecular biology. He simply did not believe that it would be possible to understand cells in molecular terms — they were, for him, impossibly complex.

One of the main conclusions of the study is how large a part 'power' plays in the scientific process, for example in what the authors call the legitimization of molecular biology. They are persuaded that Foucault's insight is "profoundly true", namely that the establishment of specific discourses — ways of speaking and thinking — also provides access to power and control over the objects of those discourses. For them "the new model that came to dominate biology after the discovery of DNA only became established after a good deal of politicking and propaganda", in which 'public relations' used by scientists had a large part. The brief analysis of this period is most unsatisfactory and fails to take into account the change in paradigm from metabolism to information. The authors also do not understand the revolution molecular biology caused, and instead concentrate on the idea that molecular biology was a philosophical movement based on the conviction that all biology could be explained in terms of physics and chemistry. Thus it is said that Crick used Schrödinger's name to legitimize his reductionist approach. Of course there was evangelizing and politicking but none of that would have been effective had molecular biology not provided fundamental new insights. Moreover, scientists are aware of these issues and the nature of scientific politics, fashions and power. That's what they often gossip about.

There is a touching, but nevertheless irritating naivety to the conception of science purveyed in the book. For example, the authors are apparently appalled by "the contradiction between the ideal norms of science, that scientific knowledge should be disinterested . . .

and the spirit of competitive individualism", and imply that competition between scientists (the "winner takes all" ethos) is something new. Yet one of the rewards in science has always been priority in discovery — are Charlesworth *et al.* really unaware of the great historical controversies, such as that between Newton and Leibniz over who discovered the calculus? And where do these 'ideals' and 'norms' come from? These really important social constructs are neither questioned nor analysed.

Competitiveness is the subject of a section of the book dealing with immunoparasitology and the quest for a malaria vaccine. Nossal's refutation of their views, which they quote at length, seems to have made no impression. Nossal recognizes competitiveness but emphasizes the tremendous amount of collaboration that science involves. "Now the question is, does the winner take all? Of course not! The peer group is not stupid. . . . The second or third in the race may not win the Nobel prize, but there are plenty of important subsidiary rewards".

This book was intended to convey a sense of the "life-world" of a group of scientists as well as some idea of how that world is "constructed". In this it fails. Much of the description is like that found in an annual report, and any half-decent journalist would have done a far better job. Rather, its value lies in what it reveals about anthropology and social science. Did the authors bring any new insights or techniques from their discipline? There is hardly a word about how the study was carried out or over what period. Did they listen to the gossip? All sorts of important questions are not touched upon. How, for example, does the Institute deal with scientists that are no longer productive; how are resources distributed; to what extent is the research directed; were there different styles of sciences; what are the rewards that scientists seek? An apparent desire for objectivity seems to have killed any curiosity, and any interest.

What, for example, can one say of the following? "What impresses me more and more about a good deal of science at the Institute is the concentration on 'getting data'" and "the idea of data generation seems to me to be a crucial feature of science and is often overlooked. The laboratory is a kind of data factory". No one should be surprised that data are important for biologists, while the metaphor of a "data factory" is plain fatuous. If this is all that anthropology has to offer for studying other groups, then the result is depressing. □

Lewis Wolpert is in the Department of Anatomy and Developmental Biology, University College and Middlesex School of Medicine, Windeyer Building, Cleveland Street, London W1P 6DB, UK.

Cut in production

Norman Myers

Tropical Forests: Botanical Dynamics, Speciation and Diversity. Edited by L. B. Holm-Nielsen, I. C. Nielsen and H. Balslev. Academic:1989. Pp.380. £29.50, \$49.50.

Tropical Rain Forest Ecosystems: Biogeographical and Ecological Studies. Ecosystems of the World 14B. Edited by H. Lieth and M. J. A. Werger. Elsevier:1989. Pp.713. Dfl.475, \$250.

The Fate of the Forest: Developers, Destroyers and Defenders of the Amazon. By S. Hecht and A. Cockburn. Verso, London:1989. Pp.256. £16.95, \$24.95.

Saving the Tropical Forests. By Judith Gradwohl and Russell Greenberg. Island Press, Washington, DC/Earthscan, London:1989. Pp.214. Hbk \$24.95; pbk £6.95.

WE ARE being overwhelmed by books on tropical forests, some of them less than welcoming. The rush to publication reflects the recent burst of interest in the forests as we become aware of their impending demise and their intrinsic wealth.

One of the reasons why we have not done a better job of safeguarding the forests is that we understand all too little of what we are dealing with — at least as compared with savannahs, grasslands, temperate woodlands and tundras — or of what is at stake. There is still only a sparse selection of solid scientific writings on the forests' makeup, their biotic richness, their ecological complexity, their community structures, their dynamic interactions and — above all — their relationships to the human enterprise. For example, we have only a vague notion of how biotically diverse the forests are. Just this year Professor Steve Hubbell of Princeton University revealed that a 50-hectare plot in Peninsular Malaysia contains over 800 species of woody plant, or as many as in all of North America. And when we finally gain a true grasp of the forests' ultimate richness, we shall finally gain a true realization of how far we have to go before we understand their diverse interactions.

Tropical Forests: Botanical Dynamics, Speciation and Diversity provides an abundance of the types of data and analyses that must underpin programmes for conservation and sustainable use. The opening section on forest dynamics appraises the physical factors such as soil and water regimes, together with patterns of tree distribution, gap-phase dynamics, canopy composition and dynamic stability of forest structures overall. There follow extended sections on speciation and diversity, primary and secondary formations, and community dynamics, before the book winds up with a retrospect-and-

prospect summation. All of the chapters are authoritative to a degree — among the contributors are Gentry, Halle, Hartshorn, Mori, Oldeman, Polhill and Raven — and they challenge the reader with the thought that tropical forests remain virtually a terra nova to modern science.

Tropical Rain Forest Ecosystems, published in Elsevier's *Ecosystems of the World* series, reflects the experiences of another lengthy list of luminaries — Bruenig, Ewel, Golley, Janzen, Jordan, Kartawinata, Kira, Prance, Sanchez and several others. The book's structure is similar to that of the first. It starts with the physical environment (notably climate among other limiting factors), moves on to tree architecture, taxonomy and classification, then to accounts of interactions such as epiphytism and symbiotic relationships, which lead into a consideration of some plant categories such as lianas, bryophytes and lichens, as well as of mammal categories such as primates, bats and flying insects. These expository reviews are interspersed with assessments of the three main tropical forest regions, and of a number of the countries concerned.

These two books make plain how few scientific answers we yet possess, and how far we are from even asking all the right questions about tropical forests. Regrettably and remarkably, however, neither of them says much about the actual rate at which the forests are disappearing, even though there is now a host of remote-sensing and other inventorying technologies that can provide the relevant data. I estimate that during the 1980s deforestation has accelerated by roughly 50 per cent, not only in Brazil but in Bolivia, the Ivory Coast, Burma, Thailand, Eastern Malaysia and the Philippines. Incidentally, the United Nations agency charged with responsibility for tropical forests, the Food and Agriculture Organization, asserts there has been no evidence of any increase in the deforestation rate during the 1980s — and it does not intend to publish a revision of its 1982 report before 1992 at the earliest.

Nor do the two books attempt any analysis of why the forests are disappearing so rapidly. The editors might legitimately claim that they are natural-science assessments. All the same, it seems strange that in 1989 they should include no multidisciplinary evaluation of the whys and wherefores of deforestation — the economic, political, social and institutional factors at work. One could argue that it is partly because of this deficiency in the past that the forests are now disappearing; ecologists and economists have been far too reluctant to make common cause in presenting an integrated assessment of what the forests amount to and hence how they can be safeguarded.

All the more welcome, then, is the third

NEW TITLES

PEST CONTROL

Second Edition

*New Studies in
Biology Series*

H F van Emden

£7.50 net approx paperback

ISBN 0 7131 2982 4

128 pages illustrated
probable publication

November 1989

BIOLOGICAL NOMEN- CLATURE

Third Edition

Charles Jeffrey

£6.95 net approx paperback

ISBN 0 7131 2983 2

96 pages probable publication
December 1989

LIGHT AND PLANT RESPONSES

A study of plant
photophysiology
and the natural
environment

T H Attridge

£9.95 net approx paperback

ISBN 0 7131 2973 5

160 pages illustrated
probable publication

February 1990

Edward Arnold
A division of Hodder & Stoughton



Felled rainforest — making a fast profit, but leaving greater wealth unrecognized.

book under review, *The Fate of the Forest: Developers, Destroyers and Defenders of the Amazon*. Confined to a single, though the largest and most controversial sector of the biome, many of the authors' insights and conclusions apply elsewhere. The true wealth of the Amazon is little appreciated by the region's political leaders; on the contrary, they often view the forest as a barrier to progress (a term they rarely if ever define). Moreover the recent outbreak of destruction is a departure only in scale. For hundreds of years there have been hands ready to wield the machete and matchbox, and to make a fast profit from just a few forest products while disregarding the many others that could offer a greater return — and do so sustainably.

Hecht and Cockburn's historical perspective is illuminating, and they are to be congratulated on their dispassionate appraisal of the forces bent on deforestation. Salvation for the forest will not depend so much on what goes on inside it, still less on the efforts of conventional foresters. It will come from the land-use planners, economists and development technocrats who operate at the macro-level of national policy, and it will reflect the determinants of agriculture, population, urbanization, financial stability, social equity and political security. These factors will have far greater effect than will logging methods, silviculture, genetic breeding and wood chemistry, or zoning systems, nature preserves and policing efforts. Most important of all, such factors will be determined in places far from the forests, even beyond the countries principally involved, for example in the citadels of foreign investment and the sources of foreign debt.

Finally, *Saving the Tropical Forests* offers a host of specific solutions to the problem declared by its title. The book contains case-studies of 38 examples of success at the local level, covering such

broad themes as natural forest management, forest restoration, sustainable agriculture and forest reserves. Among the specific topics addressed are agroforestry, home-garden systems, extractive reserves, watershed management, medicinal plants, multiple-use conservation units, buffer zones and esoteric experiments such as iguana ranching. The book is aimed at conservationists, biologists and policy makers, and presents a fine compendium of recent information on management measures and conservation strategies that demonstrably work — albeit on a small scale — in a variety of economic, sociocultural and political contexts.

But it is a commentary on these four books, excellent within their limitations though they are, and on the traditionalist approach to tropical forests, that none of them ventures a mention of what may soon prove to be the biggest factor of all in the forests' future — the climate connection. It is becoming apparent that tropical forests play a gyroscopic role in regulating climate, whether at local, national, regional or global level. The consequences could be crucial not only for the two billion people who live in the humid tropics, but also for the rest of us. Herein could lie the single most compelling reason for a 'safety first' response to all questions of forest management and conservation. Within ten years, we may well find that the most productive way to exploit the forests is to leave them to get on with their job of maintaining precipitation patterns, temperature balances and the carbon budget of the atmosphere. □

Norman Myers is at Upper Meadow, Old Road, Headington, Oxford OX3 8SZ, UK.

● Recently published by Barrie & Jenkins, London, *The Rainforests: A Celebration*, edited by L. Silcock, is a superbly illustrated book of the rainforests' beauty and variety, price is £19.95.

Predictable outcome

Philip Gummatt

Research Foresight: Priority-Setting in Science. By Ben R. Martin and John Irvine. Pinter, London:1989. Pp.366. £35, \$45.

In his *New Atlantis*, Francis Bacon describes the order or society called Salomon's House. This seventeenth-century prototypical research organization operated a well-planned division of labour. Some of the fellows surveyed the literature; others devised or performed experiments; yet others explored the theoretical implications of research results; and some (called 'Benefactors') sought applications of this work. Another important role was performed by those who travelled abroad, in disguise, to collect the books and experiments of other countries. These were called Merchants of Light.

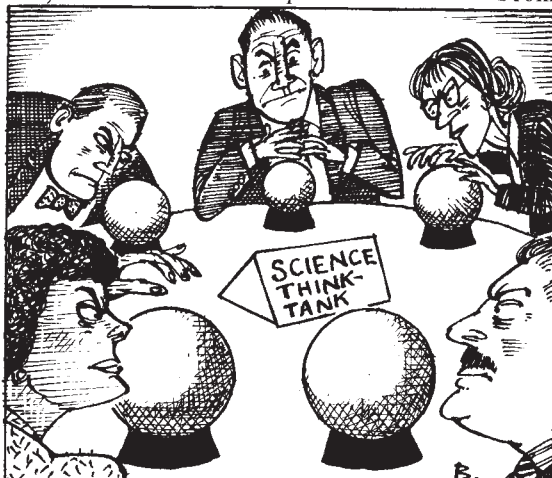
Today, when the state demands increasing returns from its investment in science (so that the 'Benefactors' might be renamed 'Exploiters'), Bacon might think it wise to add a cadre of policy analysts. They would need their own Merchants of Light, for science policy is now an international business. Two candidates might be the authors of the book under review, who have travelled the world, in the guise of contractors to the Dutch government, bringing back information on how eight countries engage in 'foresight' in science and technology. Oddly, having gone to France, West Germany, the United States, Japan, Australia, Canada, Sweden and Norway, they omit their own country, Britain. Their excuse (lack of time and difficulties over commenting on foresight exercises in which they had been involved) is not entirely compelling.

'Foresight' is a slightly curious term. For one thing, what should we call the people who do it? 'Foreseers' sounds too biblical, while 'forecasters' misses a crucial distinction. As the authors explain, the aim is not to forecast, nor — as was suggested by the subtitle of their previous book, *Foresight in Science: Picking the Winners*, published by Pinter in 1984 — to pick winners. Rather, it is to establish a continuous process for reaching an improved understanding of possible developments in science and technology, and the forces likely to shape them, in an attempt to nurture winners and hence create the future.

In their travels, Martin and Irvine encountered a great deal of activity of these types, especially in relation to strategic and applied research (less so for basic research). There were, though, considerable variations between countries in terms of methods and institutions used,

and commitment to the foresight process.

Some countries use formal bibliometric methods to try to spot growth points in science, nationally and internationally. Some engage in elaborate consultative exercises, involving large numbers of scientists and users of science and technology. Some adapt various tools from the kitbag of forecasters and futurologists. Some set up specialist panels on narrow subjects, while others try to take a view of the whole national scene. Some draw on professional analysts (especially Japan, which has a network of some 300 public and private think-tanks, many of them with levels of expertise that are unmatched even by the large US consultancies). Some use what a respondent in the



United States called the BOGSAT method (Bunch of Guys Sitting Around a Table).

What lessons can be extracted from this bewildering variety? The Director-General for Science Policy of the Dutch Ministry of Education and Science, in introducing the book, singles out three. First, analysis of long-term needs and opportunities for research should be part of a well-planned process that includes implementation of the choices made. The authors identify a number of attempts at foresight that failed because it was not clear from the outset what was to be done with the results. Second, it is essential, especially for smaller countries, to carry out systematic global monitoring of developments in research. This is something that the Japanese have been doing since the 1950s. Third, and as already indicated above, the aim of policy should be to nurture scientific winners rather than simply to try to pick them, ready-made as it were. That is, a process of identifying, and then supporting, areas deemed to be of national importance should be established.

Other lessons include, first, the point that introducing foresight approaches can be a slow and painful process, especially in countries with no culture of planning or consensus-seeking. This is particularly so in relation to setting priorities in basic

research, even in Japan. Second, the authors show the importance of trying to integrate the different levels of the policy process, from all-embracing national socio-economic goals, down through various sectoral institutions and objectives, to strategic and basic research. But they also show the crucial importance, in seeking this integration, of achieving a balance between the views of innovative young scientists, who can see the possible directions in which science might go, and the more seasoned views of senior officials, industrialists and other users, who can relate science-push to social demand, particularly across the boundaries of established disciplines.

From a parochial angle, it is a pity that Britain was excluded from the study. Britain is a country in which formal foresight methods of a bibliometric type have been used by, for example, the Royal Society and the Advisory Board for the Research Councils. The Cabinet's Advisory Council on Science and Technology, and its predecessor, have conducted a number of studies with the intention of guiding policy over research priorities. The University Grants Committee, now the Universities Funding Council, has engaged in substantial restructuring of basic research, though without (as far as one can tell) any vision of the national future that it is aiming to create. And, in response to the arguments in Irvine and Martin's 1984 book, the Centre for Exploitation of Science and Technology — which is mainly funded by industry — has been set up to try to articulate industrial views of the future and integrate them with government policy. The views of our Merchants of Light on these measures would have been welcome.

We can be sure that foresight is here to stay (which may make it worth trying to find a better name for it). The perception is growing that the procedure does not imply draconian planning. Instead it aims at improving communications between producers and users of research and arriving at decisions based on a consensus about the choices to be made. The winners, so the advocates of foresight claim, will be the countries that manage this process best. □

Philip Gummatt chairs the Board of Science and Technology Policy, University of Manchester, Manchester M13 9PL, UK.

● Recently published by Routledge, London, is *Innovation and Technology Transfer in Japan and Europe: Industry-Academic Interactions*, by Glyn O. Phillips. The book provides "a perspective against which we [the British] can take decisions about the future of our own industry — education interaction arrangements" and costs £45.

Fact and fiction

S.A. Barnett

The Longman Literary Companion to Science. Edited by Walter Gratzer. Longman: 1989. Pp. 517. £17.95. To be published in the United States by Norton.

From Creation to Chaos: Classic Writings in Science. Edited by Bernard Dixon. Basil Blackwell: 1989. Pp. 280. £15, \$29.95.

HERE are two compendia about aspects of the scientific life. Both, especially Gratzer's, can be read with enjoyment; both are highly misleading.

Gratzer wishes to "explore the incursions of science into literature". Happily, his 216 excerpts (each of which has a brief editorial introduction) are not only from stories and poetry; many are highly readable and informative passages from biography and journalism. Readers should not be deterred by the unrepresentative false start. The first section, "The Laboratory Scene", opens with a series of exposés of "the research supervisor as a monstrous incubus". Supervisors are far more often harassed than depraved. In default of a quantitative survey of supervision, we need a David Lodge to write a hilarious but authentic novel on the struggles of a supervisor to get her pupils to write up their work, and her final resort to writing the papers herself.

After the supervisor, we have "the research grandee and the array of his vassals". Nobel prizes, and their winners, though rare in real laboratories, loom large in this collection. Only by the fifteenth entry do we reach one of the few fictional descriptions of what research is really like: in Nigel Balchin's *A Sort of Traitors*, two research assistants patiently but resentfully record many thousands of cultures as showing "slight growth". And, much later, there is a comment on Ernest Rutherford (in Gratzer's words, "probably the greatest experimental physicist of this century"): "he got pleasure not only from the high moments, but also from the hours of what to others would be drudgery, sitting in the dark counting the alpha particle scintillations on the screen". Another rarity is a poet, Louis MacNeice, with his unromantic but not unkind portrait of a scientist who

*... lived by measuring things
And died like a recurring decimal
Run off the page, refusing to be
curtailed;
Died as they say in harness, still
believing
In science, reason, progress.*

A book such as this could hardly dwell at length on quiet people and non-events. It is appropriate that we should be horrified by the effects of anti-semitism in

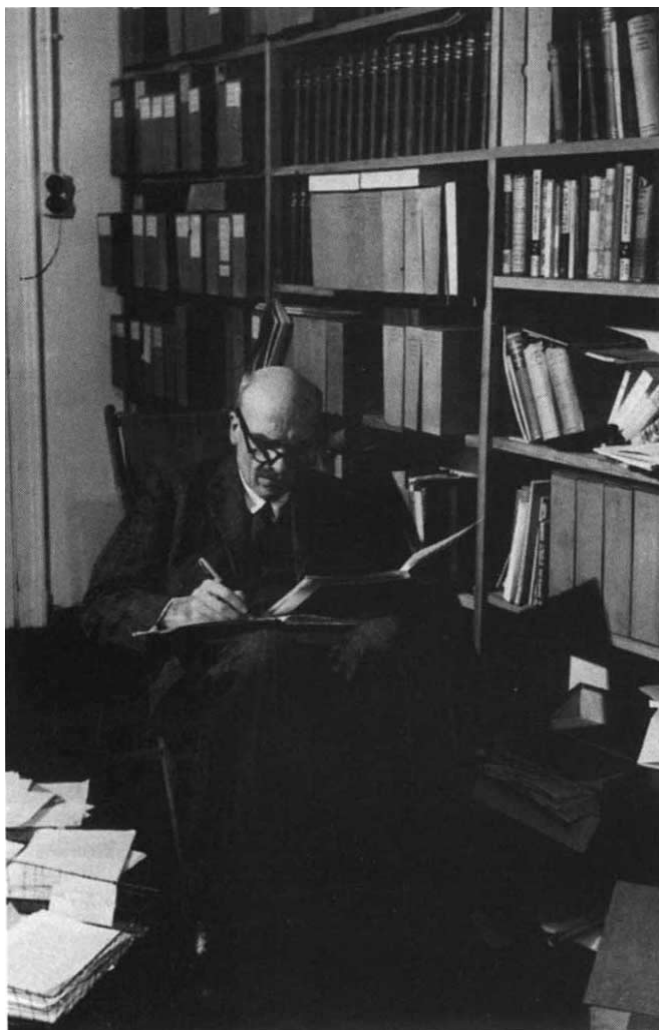
Poland, or of McCarthyism in the United States; that we should be frightened by an account of accidental death due to radiation; and sickened by the outcome of mindless bombing in the Second World War. In contrast, we may applaud the case of the brilliant but wayward student, admitted to a course of study though without qualifications. And we may enjoy passages from those masters of satire, Jonathan Swift and Samuel Butler, or turn for light relief to stories of the great in disconcerting situations — for instance, Albert Einstein trying to steal tobacco from Niels Bohr.

Gratzer's exertions do not seem to have been entirely a labour of love: he comes over as ambivalent towards scientists. He quotes a tribute by a classical scholar, Maurice Bowra, who said that scientists were treacherous allies on committees because "they are apt to change their minds in response to arguments", but he immediately qualifies it by reference to scientists' "passions" and, in a section on "Discovery", he states that "emulation and jealousy among scientists have [today] become sanctified as the motives that drive scientists ever onwards". Here, perhaps, is another hypothesis in need of a survey.

He finishes, as he begins, with bathos: the last section, "The Pathology of Science", contains a mixed bag of accounts of fraud and plagiarism by scientists and bogus science by illusionists, with only one (fictional) example of commercial skulduggery. Gratzer could have acknowledged that his patchwork is far from a representational portrait of science by ending with a remark by the physicist, Jeremy Bernstein, quoted on p. 151: "I pity the historians of science and others who are attempting to put together a reliable history [of science]; all you have to do, you may think, is ask the participants for their accounts of what happened. But such stories can be woefully misleading".

Bernard Dixon, too, is concerned with literature: by publishing his collection, he hopes to "demolish the well-worn cliché

that scientists, by their nature and training, cannot write", although his medley of excerpts includes many by science writers (and even a rather dim limerick by Arthur Koestler). Most of his preface is directed at encouraging young scientists to write well. He warns adolescents against trying to imitate "models of unattainable excellence", and says that instead they should learn by analysing bad writing and read great prose for pleasure so that "lessons of style... sink in by subconscious osmosis". But he ignores the principal, perhaps the only, effective method of developing order, clarity and conciseness, which is to practise and to subject one's



Capable writer — J.B.S. Haldane at University College, London 1948.

attempts to criticism.

Dixon also tells us that his book's "sole *raison d'être* is to provide pleasure for the reader". Except that they are said to be selected "solely on the basis of literary quality" the writings are unfortunately presented without editorial comment. Yet the pleasure and instruction one derives from reading are due to the combined effects, often inextricable, of style, content and context.

After the confusing preface, a reader should not be surprised to find a rather uneven selection. The passages are

arranged alphabetically, hence are randomized. Six items are in verse. Of 69 prose excerpts, 18 are fragments that convey little and hardly do justice to their authors. They could well have been replaced by longer passages from writers such as Bertrand Russell and Joseph Needham who are not represented.

Among the remainder are several which, I hope, will send readers back to the originals. Here is Haldane, in 1932, on the evolutionary process: "In the first place, it is very beautiful. In that beauty is an element of tragedy. . . . In an evolutionary line rising from simplicity to complexity, then often falling back to an apparently primitive condition before its end, we perceive an artistic unity similar to that of a fugue". And here is Peter Medawar (1960), another polymath, writing less agreeably on the same subject: "It is a profound truth . . . that nature does *not* know best; that genetical evolution, if we choose to look at it liverishly instead of with fatuous good humour, is a story of waste, makeshift, compromise and blunder".

Human evolution appears in questions asked by A.R. Wallace (1875) in an unfamiliar idiom: "whether we compare the savage with the higher developments of man, or with the brutes around him, we are alike driven to the conclusion that in his large and well developed brain he possesses an organ quite disproportionate to his actual requirements — an organ that seems prepared in advance, only to be fully utilized as he progresses in civilization".

Wallace also appears in an extract from Darwin's *Autobiography* of 1876, where the author scolds himself for overlooking (in the 1840s) an important problem — "the tendency in organic lines descended from the same stock to diverge in character as they become modified. The solution, as I believe, is that the modified offspring of all dominant and increasing forms tend to become adapted to many and highly diversified places in the economy of nature". An alternative autobiographical model is provided by Otto Frisch (1979), who describes a visit to his aunt, Lise Meitner, in exile in Sweden, which led to a major discovery in atomic physics. Or one may turn to H.G. Wells (1934) on his unsuccessful attempts, as a hamfisted student in the 1890s, to make apparatus, and his all too successful campaign to burlesque his professor; or to Miriam Rothschild writing cheerfully, in 1987, on having seven years of work destroyed by a Nazi bomb.

The reader may also compare T.H. Huxley's famous account of a piece of chalk (1868) with Primo Levi's still more readable story of a carbon atom (1984). Rachel Carson's brilliant *Silent Spring* (1962) may be contrasted with E.J.H. Corner's lucid *Life of Plants* (1964). A

problem outside the natural sciences is presented by S.J. Gould (1984) in a discussion of Adam's navel — an enigma indeed, for Adam was created, not born. These and other passages show that a few people with scientific training write well, even superbly. Classified and put briefly in context, they could have made a treasury to be read both for pleasure and for

instruction. I especially emphasize the second purpose, for such a volume is needed to help in remedying a lamentable state of affairs — that most scientists still write badly. □

S.A. Barnett is Emeritus Professor of Zoology in the Australian National University, Box 4, GPO, Canberra, ACT 2601, Australia.

Blaming biology?

Daniel J. Kevles

Dangerous Diagnostics: The Social Power of Biological Information. By Dorothy Nelkin and Laurence Tancredi. Basic Books: 1989. Pp. 207. \$18.95.

BIOMEDICAL research, especially in genetics and the neurosciences, has been generating an increasing range of new diagnostic methods and machines. Medical genetics can specify several thousand single-gene diseases and disorders, while magnetic resonance imaging, PET-scan and other techniques can detect numerous abnormalities in the brain. The new diagnostics reaches beyond the detection of physical maladies such as Tay-Sachs disease or brain tumours to identify tendencies to particular types of behaviour or mentality, drawing on recent scientific research that has found, for example, genetic sources for schizophrenia or brain lesions linked to learning disabilities. To be sure, the growth in diagnostic power has far outpaced therapeutic capacity. But never mind: diagnosis has its uses independent of cure. Diagnosis produces information about individuals that can assist life decisions they make for themselves — or that are made for them by others.

Genetic assessments have for some years figured in clinical settings, typically permitting prospective parents to learn the risk of conceiving a child that suffers from a genetic disease or assisting the physician in identifying the physical cause of a patient's malady. However, the broadening powers of neurophysiological and genetic tests are making them evermore appealing as devices to screen individuals with benefit, it is said enthusiastically, not only to themselves but to society. The National Institute of Mental Health has expressed a commonplace expectation: "Pre-symptomatic detection of psychiatric disease will be routine. Subjects at risk for alcoholism, schizophrenia, or depression may be identifiable well before the onset of clinical symptoms" (pp. 5–6). Families and physicians will thus be better prepared to take preventive action or apply therapeutic measures.

Diagnosis is similarly touted for use elsewhere in society. In the schools, it will

assist teachers in dealing with troublesome children; in the courts, help specify the degree of a defendant's mental capacity; in the insurance world, aid in the determination of a client's health risk and, hence, premiums; and in the workplace, identify the propensity of workers to fall ill in response to, say, chemicals. According to the *Journal of Occupational Medicine*, diagnostic screening makes good business sense because it can "lead to reduced absenteeism, increased productivity, and decreased expenditures for workers' compensation and group health insurance" (p. 90).

In *Dangerous Diagnostics*, a critical survey and analysis of the diagnostic trend, Nelkin and Tancredi allow that the testing may indeed be valuable — as is, for example, DNA fingerprinting in the courts — but they argue, with arresting persuasiveness, that much of it is alarming. Nelkin, a sociologist at New York University, is an authority on the affairs of science and society. Tancredi directs the Health Law Program at the University of Texas Health Science Center in Houston. Their respective talents and expertise are advantageously joined in this book, which they wrote to stimulate debate about "the potential uses and abuses" of the emerging biological tests and as a kind of technology assessment addressed to their inherent social and political implications (p.ix).

Among the strengths of Nelkin and Tancredi's treatment is their insistence that the danger in diagnostics arises not so much in any evil intent as in the way that institutional contexts, and the values inherent in them, encourage the assessments to be used — and abused. A rapidly growing diagnostics industry presses the tests upon the world ("The future lies in genetic tendencies", an industrial biotechnologist has declared — p. 5). In the courts, reliance on 'hard' biological evidence of mental condition may improve the efficiency and reduce the costs of the judicial process; it can also exclude traditional psychiatric testimony that places the mental state of defendants, to their advantage, in a broader setting. Diagnostic screening in the workplace may improve productivity and decrease liability; it also shifts responsibility for workplace hazards from the company to the employee. Screening may better reveal an individual's risk of disease; but it enables health-care providers and

insurance companies better to discriminate against potentially costly clients.

The dangers are amplified by an uncritical embrace of the science on which some of the diagnostics rest. Employers have been known to screen job applicants for characteristics — sickle-cell trait or pregnancy, for example — that are irrelevant to hazards in the employers' workplace. Depending upon the test, screening of a random population can yield a comparatively large number of false positives. Identification of a genetic risk for a disease such as manic depression is often taken to mean a genetic certainty to succumb; similarly with aberrant brain patterns. In both cases, the complex etiology of many diseases is ignored in favour of single-cause assessment and economic, social or medical stigmatization.

In the view of Nelkin and Tancredi, the diagnostic dangers revive an old theme — a propensity to medicalize socioeconomic misfunction, to locate its principal causes in purely biological traits. In the ancient debate between nature and nurture, nature, they say, has recently been declared the winner. Blaming biology lets society or its constituent institutions off the hook; it permits them to segregate problems rather than deal with them. It fosters the construction of social categories — 'high-risk' employees or 'learning-disabled' children — and places them beyond the reach of meliorative social policy. Nelkin and Tancredi write, "By focusing attention on the physical state of the brain, educators may neglect the social and economic conditions leading to damage. By dealing with the biological characteristics of individuals, rather than the social causes of problems, schools set up solutions that insulate them from responsibility while enhancing their immediate control" (p.130).

Yet it should be remembered that diagnostics need not result in a deterministic medicalization. Identification of a biological handicap can make possible compensatory environmental remedy. It can also help society more knowledgeably define where it wishes to set the limits — do haemophiliacs have an inalienable right to employment as butchers? — of regulatory or environmental accommodation. One wishes that Nelkin and Tancredi had given more attention to the technical quality of the new diagnostics, especially of the neurophysiological variety: technical reliability must inevitably figure in assessments of their appropriateness. One wishes, too, that they had explored more extensively the genuine challenges and trade-offs with which the new diagnostics confront society; for example, how the costs of risk should be allocated as risk becomes more sharply specified; or how institutional interests should be balanced against individual rights, particularly rights to education,

employment, health care or privacy.

Not least because it points to the need for such explorations, *Dangerous Diagnostics* is an important and timely book, a thoughtful and informative provocation to debate. It is also a pathbreaking one, for its recognition that while the diagnostics may be applied across a broad informational and institutional spectrum, their increasing uses raise a class of social and political issues that are general, and generally disturbing. For as Nelkin and

Tancredi apprehensively declare, "If biological tests are used to conform people to rigid institutional norms, we risk reducing social tolerance for the variation in human experience. We risk increasing the number of people defined as unemployable, untrainable, or uninsurable. We risk creating a biologic underclass" (pp. 175–76). □

Daniel J. Kevles is in the Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, California 91125, USA.

The price of sex

Rory Howlett

Sexual Selection. By James L. Gould and Carol Grant Gould. W.H. Freeman: 1989. Pp. 277. £16.95, \$32.95.

SEX is costly and in many ways disadvantageous, yet almost all species alive today reproduce sexually. Others, however, are either asexual or alternate between sexual and asexual phases, suggesting that there is a fine balance between the costs and benefits of sex.

The subject is a controversial one, attended by a host of unanswered questions. In this addition to the Scientific American Library, James and Carol Gould set out to explain the roles that sex (technically, the recombination of separate sets of genes) and mate-acquisition

centrates on the ways in which selection has affected the lives of sexually reproducing species, from non-social insects to the most social of all species, us. Here there are entertaining accounts of how animals choose and woo their mates, and how they use even the most devious of means to ensure that their genes are propagated in future generations; indeed, as has become almost obligatory since the publication of *The Selfish Gene*, a gene's eye view is taken throughout.

As the Goulds point out, the special challenge of behavioural ecology is to explain why there are so many different social organizations in the animal world, and why each species structures itself in the way it does. A particularly satisfying aspect of the book is the way that different social organizations are compared and related to the differences in habitat and niche; resource distribution and availability have a central role in determining the level of sociality that evolves.

Even within a species, individuals often vary in their behaviour, and it was previously believed that minority behaviours were somehow deviant. The authors do a fine job of explaining how frequency-dependent selection can give rise to a stable mixture of 'strategies' within species, an idea that has derived in large part from game theory. Equally well covered is the contentious topic of the evolution of female choice and its role in selecting for apparently disadvantageous (in terms of natural selection) male secondary sexual characters such as the tail of the peacock. Finally, we have a thought-provoking (and unashamedly 'sociobiological') account of the role of sexual selection in shaping human evolution and culture.

The Goulds write clearly yet with authority (both are involved in research on sexual selection) and lace the text with many fascinating facts and examples. Their book will do nicely for students and others with an interest in the subject. □

Rory Howlett is an Assistant Editor of *Nature*



Like the Red Queen — running as fast as you can to keep up.

strategies have had in the lives of organisms from bacteria to human beings, and to tackle some of those many questions.

The Goulds start by summarizing the origin and maintenance of sex, placing it squarely in the context of modern evolutionary thinking. Working on the assumption that there must be a powerful logic to explain the near ubiquity of sex, they compare and contrast an "embarrassment" of plausible hypotheses. Their conclusion is that, at the outset, the role of sex in genome repair was probably of overriding importance, while selection in response to ever-changing challenges (red-queen evolution) is likely to be the main factor maintaining sex in most modern species. The rest of the book con-

A strange type of addiction

David L. Hull

The Selfish Gene, new edition. By Richard Dawkins. Oxford University Press: 1989. Pp. 352. Hbk £17.50, \$22.95; pbk £5.95.

The Selfish Gene appeared in 1976. Dawkins wrote the book to educate general readers about biological evolution, to prod experts out of their slumbers over the issue of individual selection and to catch the imagination of students. He succeeded on all counts. For this second edition, he has not engaged in rewriting on a large scale. Instead he has added two chapters at the end and interspersed his text with footnotes in which he acknowledges errors, expands on certain points and replies to his critics. The result is a more moderate and conciliatory book than the original. Dawkins has caught our attention. Now he can afford to present his views in more measured terms.

One of the issues that has fascinated Dawkins is exactly how nice we can be given the constraints placed on us by biological evolution, in particular the 'selfishness' of genes. He is not impressed by objections that genes can no more be selfish or unselfish than atoms can be jealous. Scientists have always borrowed words from ordinary language and transformed them into technical scientific terms. If Newton could use the Latin equivalent to 'force' to characterize mass, and quantum physicists can talk about a subatomic particle having 'charm', then biologists are justified in turning 'selfishness' into a technical term as well. Unfortunately, some choices can lead to misunderstandings — the selfishness of genes in the technical sense has consequences for the selfishness of organisms in a more ordinary sense, and the two are easily confused.

Dawkins is swayed more by objections that turn on the social implications of his way of expressing himself. He began working on *The Selfish Gene* in 1972 when power-cuts resulting from industrial strife forced him to interrupt his laboratory research. At the time the behaviour of the working classes in Britain seemed to offer increasingly good examples of selfish short-sightedness. But times have changed: "Now that Britain has a government of the new right, which has elevated meanness and selfishness to the status of ideology, my words seem to have acquired a kind of nastiness by association that I regret" (p. 268). He also regrets his overly cynical and selfish view of life expressed in his chapter on the battle of the sexes, but at the time he felt the need to counteract the Pollyannaish view then prevalent of the relation between the sexes.

In the original book, Dawkins acknowledged that a "human society based simply on the gene's law of universal ruthless

selfishness would be a very nasty society in which to live". In part he intended to warn us that if we want, as he does, to "build a society in which individuals cooperate generously and unselfishly towards a common good, you can expect little help from biological nature" (p. 3). Our only hope for a better society is to understand our biological make-up and to counteract it. As surprising as it might sound coming



Dawkins — catching our attention. from someone who has so strongly emphasized the role of genes in biological evolution, Dawkins nevertheless insists that "Darwinism is too big a theory to be confined to the narrow context of the gene" (p. 191). Genes function as the replicators in biological evolution while memes serve this same function in cultural evolution.

Early darwinians such as Huxley presented their contemporaries with the choice of either darwinism or miracles. Dawkins seems to be presented with a comparable dilemma — either genetic determinism or free will. He insists that he was not and never has been a genetic determinist. Genes exert the same sort of statistical influence on behaviour as on other phenotypic traits. Genes are necessary. They are never sufficient. The influence of genes can be "modified, overridden or reversed by other influences" (p. 331). These other influences are to be found in our environment, especially in our society. "We, that is our brains, are separate and independent enough from our genes to rebel against them" (p. 332). The crucial issue is our power to influence which memes come to reside in our brains. Memetic determinism is no less deterministic than genetic determinism. We can rebel against our genes, but can we rebel

against our memes? Science is one way. If science is anything, it is a method that can force us to change even our most cherished memes.

The two additional chapters added by Dawkins are entitled "Nice Guys Finish First" and "The Long Reach of the Gene". The first corrects and expands upon his earlier discussion of evolutionarily stable strategies. If one looks at simple versions of such game — theoretical problems as the Prisoners' Dilemma, it is difficult to see how cooperative behaviour can become prevalent. Nice guys cannot win. But in real life organisms play repeated (or iterated) games. Behaviour that has a low payoff if each game is made up of new players can become prevalent if successive games are played with the same contestants. In iterated games it turns out that the nastier strategies rarely prevail — and selection is inherently an iterated game. Thus, Dawkins concludes that "even with selfish genes at the helm, nice guys can finish first" (p. 223). But notice that genes are still at the helm.

The other new chapter is not concerned with who is nasty and who is nice but with the basic structure of selection processes themselves. For all the modifications that Dawkins has introduced into his theory, he still argues for the primacy of genes, but he now sees more clearly in what ways they are more primary than organisms. In selection processes entities must replicate themselves. They must interact causally with their environments in such a way that replication is differential. Differential replication in the absence of such environmental interaction is drift, not selection. Dawkins terms the entities that perform the first function "replicators". According to him, genes are the sole replicators in biological evolution. Replicators can interact directly with their environments or produce more inclusive entities (vehicles) that perform this function for them.

One of the reasons why Dawkins wrote *The Selfish Gene* was to combat the alacrity with which so many ethologists explained certain behaviours in terms of group selection. "We can now see that the organism and the group of organisms are true rivals for the vehicle role in the story, but neither of them is even a *candidate* for the replicator role" (p. 254). Organisms and even certain sorts of colonies can function in selection processes — as vehicles, not replicators. Dawkins is a genic replicationist, not a genic selectionist.

The part of the second edition that I found most valuable is Dawkins's reworking of the foundations of our understanding of the evolutionary process. All that is necessary for selection to occur is that genes have phenotypic effects. But, in point of fact, genes are organized into genomes, they gang up in cells, cells gang up into multicellular organisms and some organisms gang up in colonies. If discrete,

Selected titles from Kluwer

Alternative Life-History Styles of Animalsedited by **Michael N. Bruton**

PERSPECTIVES IN VERTEBRATE SCIENCE 6

Fresh and sometimes controversial insights are provided on the consequences of bifurcation for epigenesis and evolution, and a suggestion is made that new paradigms are emerging which emphasize the harmonious self-organization of natural systems (cells, organisms, communities and ecosystems).

1989, 624 pp. ISBN 90-6193-662-4
Hardbound Dfl. 400.00/£129.00/US\$215.00

Inland Waters of Southern Africa: An Ecological Perspectiveby **B. R. Allanson, R. C. Hart, J. H. O'Keeffe and R. D. Roberts**

MONOGRAPHIAE BIOLOGICAE 64

The book commences with a scene-setting review (in four chapters) of the relevant aspects of the sub-continent, followed by a discussion of the river and its catchment, floodplains and wetlands, and of natural and man-made lakes. The various data are drawn together in the last section, which is devoted to a synthesis and the implications of the findings. Extensive references and index complete the work.

1989, 464 pp. ISBN 0-7923-0266-4
Hardbound Dfl. 300.00/£100.00/US\$159.00

Acid Waters in Walesby **R. W. Edwards, A. S. Gee and J. H. Stoner**

MONOGRAPHIAE BIOLOGICAE 66

The book deals with: Evidence for Acidification (Paleo-ecological and contemporary); Causes of Acidification; Effects of Acidification (on ecology, fisheries and water-supply); Catchment studies (including hydrochemical and biological modelling) and Potential Solutions. Chapters on Ecotoxicological Studies and the Chemistry of Aluminium are included.

1990 ISBN 0-7923-0493-4
Hardbound price to be announced

Evolution of Laurussia**A Study in Late Palaeozoic Plate Tectonics**by **Peter A. Ziegler**

This volume provides earth scientists with an overview of the late Palaeozoic evolution of North America, Europe, the Arctic and North Africa in a plate tectonic framework.

1990, 100 pp. ISBN 0-7923-0428-4
Hardbound Dfl. 145.00/£49.00/US\$69.00

**KLUWER
ACADEMIC
PUBLISHERS**

P.O. Box 322, 3300 AH Dordrecht,
The Netherlands
P.O. Box 358, Accord Station, Hingham,
MA 02018-0358, U.S.A.

Reader Service No.26

320

individually purposeful vehicles are not necessary for selection, why are there so many of them? And why do multicellular organisms adopt a 'bottlenecked' life cycle?

The crucial distinction for Dawkins is not between sexual and asexual reproduction, but between those organisms that periodically go through cellular bottlenecks and those that do not. In the second case, no difference exists between growth and reproduction. Mutated cells find themselves clumped with distantly related cells that retain the original genetic make-up. But organisms that go through single-celled bottlenecks can go back to the drawing board. Instead of attempting to reorganize an entire organism, all that has to be modified is a single genome. Sequestering certain cells into germ lines facilitates the evolution of complex organisms. This mode of organization seems so advantageous, one wonders why it is not universal. Clonal organisms must be doing something right because they are still with us.

Although Dawkins grants that bare genes are invisible to natural selection, he

argues that natural selection acts on genes through differences in their phenotypic effects, and these effects need not stop at the boundaries of our bodies. They extend to all the effects that a gene has on the world, including responses in the behaviour of other organisms. Thus, both Dawkins's book and the responses to it count as part of his extended phenotype. In one place, he describes an adult dunnock feeding a cuckoo chick that it has been tricked into adopting. The dunnock is so much smaller than the cuckoo chick that it must perch on the back of this monstrous foster-child to feed it. Dawkins remarks that German ornithologists refer to such foster-parents as 'addicts' and their cuckoo nestlings as their 'vice'. As I read this description, I pictured him perched atop his theory attempting to feed its gaping mouth. Scientists give every appearance of being addicts, and science is their vice. That is one reason why progress in science is so rapid. I for one have benefited a great deal from Dawkins's addiction. □

David L. Hull is in the Department of Philosophy, Northwestern University, Evanston, Illinois 60208, USA.

Facts of life

W. F. Bynum

The Timetables of Science. A Chronology of the Most Important People and Events in the History of Science.

By Alexander Helleman and Bryan Bunch. *Simon & Schuster, New York/Sidgwick & Jackson, London: 1988/89. Pp. 656, \$29.95, £19.95.*

Asimov's Chronology of Science and Discovery. By Isaac Asimov. *Harper & Row, New York: 1989. Pp. 707. \$29.95, £22.95.*

WHAT are we to make of the almost simultaneous publication of two chronological guides to the history of science and technology, especially when a third — *Breakthroughs: A Chronology of Great Achievements in Science and Mathematics* by Claire L. Parkinson (Mansell, 1985) — has already been remaindered? One conclusion is simply that publishers think such books will sell (despite the remaindered example). More significantly, perhaps, is the fact that scientists do think about the history of their endeavours chronologically. As popularly conceived, scientific knowledge is cumulative, in ways that make it matter that Newton came after Galileo, or Dalton followed Lavoisier. Dates also count in priority disputes, to say nothing of the excuse anniversaries provide for holding jamborees. Who knows, anniversarial studies may even be recognized soon as an important branch of the history of science.

Some of the attractions and pitfalls of this emerging discipline are amply illustrated by the two books under review. Asimov's is the easier to read, because he writes short, discursive essays, mostly of around 200 words, for most of the discoveries or inventions covered. There is something for every year since 1793 and for most years from about 1600. It all becomes rather formulaic: 1902 and 1970 were good years, with eleven entries each, whereas 1958 and 1928 were rather ordinary, with only five. Fleming's discovery of penicillin gets about as much space as Marguerite Perey's discovery of francium, Diderot's *Encyclopédie* about as much as the fifth satellite of Jupiter (Amalthea).

Asimov ends with the greenhouse effect, 4,001,988 years after his first entry, on bipedality, thereby giving a symmetry to the volume — if the first process had not occurred, neither would the last. In between there is much to admire but also everything to take on trust. As befits the author of 425 other books, Asimov has eschewed preface, acknowledgements, references or bibliography. The only personal touch is a dedication.

Helleman and Bunch (both of whom are science editors) don't tell us where they got their information from either, although in the preface they acknowledge help and provide the useful information that they "have relied upon the consensus of many historians of science whose works we consulted to help locate the events as closely as possible". They record about 10,000 separate items, ranging from a birth or a death to a publication, experiment or invention. Throughout, a 'general'

column locates the ordinary events of history (beginning of the English Civil War) or 'non-scientific' aspects of their story (publication of Freud's *Jokes and Their Relation to the Unconscious*, introduction of Oreo cookies by the National Biscuit Company). Their timespan is shorter than Asimov's: they also end in 1988 but don't begin until 2,400,000 BC. The spare, atomistic approach to separate bits of information makes the book impossible simply to read. But a series of nine overviews (for example "Science after World War II") and more than 100 short essays provide some social and intellectual context for entries such as "A. Meissner invents a radio transmitter with vacuum tubes" (1913), "Permanent press clothing is introduced" (1964) or "Ernst Chladni invents the euphonium, a musical instrument in the tuba family" (1789).

The assiduous reader of Asimov would not have picked up any of these latter facts. For Asimov, the key dates for the radio are 1895 (radio antennae), 1901 (Marconi's sending of radio signals from England to Newfoundland), 1904 (J.A. Fleming's rectifier), 1906 (R.A. Fessenden's work on amplitude modulation), 1916 (E.H. Armstrong's superheterodyne receiver), 1939 (Armstrong's perfection of frequency modulation), 1948 (transistors) and 1954 (improvements in the transistor). For Hellemans and Bunch, there are at least 21 key moments, in addition to Meissner's work. For them, the radio antenna did not properly appear until 1905, Armstrong's superheterodyne receiver came in 1918 and Armstrong perfected frequency modulation radio in 1933. These authors agree with Asimov on the other dates, they also add a number of other moments in the history of radio.

Discrepancies also emerge in cross-checking through a year rather than a subject. Take 1922, for instance. Asimov writes of Leonard Woolley's excavation of Ur (dated to 1918 in Hellemans and Bunch, where his name is spelled Wooley); Howard Carter's discovery of Tutankhamen's tomb (confirmed by H and B); Herbert Evans's work on what became known as vitamin E (confirmed by H and B, although with the addition of K. J. Scott); Evans's research on growth hormone (not mentioned at all in H and B); Alexander Fleming's isolation of lysozyme (dated to 1921 in H and B); Oparin's work on the origin of life (under 1936 in H and B); experiments of J. Erlanger and H.S. Glasser on nerve conduction (recorded by H and B under 1944, when the Nobel prize was awarded); and A.A. Friedmann's work on the expanding Universe (confirmed by H and B).

Who would want to be an anniversarial scientist when there is so much uncertainty about the facts of history, to say nothing of their interpretation? Not I, which is why I have steadfastly refused to

adjudicate between Asimov and Hellemans and Bunch. It would, after all, simply be my reference books against theirs, and since they have declined to reveal theirs, it would be an unequal contest. I have reason to believe, however, that Leonard Woolley spelled his name thus. □

W. F. Bynum is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BP, UK.

Of thoughts past

Graham Ritchie

Ancient Britons and the Antiquarian Imagination: Ideas from the Renaissance to the Regency. By Stuart Piggott. *Thames & Hudson: 1989. Pp. 175. £14.95, \$19.95.*

As excavation techniques extend the range of what can be recovered, and theoretical considerations crowd into university curricula, archaeology increasingly turns to science. How important it is, therefore, that in the history of ideas about early British antiquity we should be aware of the attitudes of early scholars and have an appreciation of the framework within which their thinking took place. Stuart Piggott's perceptive exploration of antiquarian ideas draws together many

their own views of the world. It is a measure of Piggott's success that within this rigorous approach the reader is so pleasurably taken on a voyage of exploration through the worlds of the computation of Biblical chronologies, ancient giants and what happened after the Flood.

The Indians and Eskimos — even a Brazilian king — who were transported from the New World to be paraded before the European public in the course of the sixteenth century had a considerable effect on antiquarian approaches and on the way in which early Britons were depicted. It is startling to realize that at least a thousand Red Indians were brought back to Europe by early voyagers and that Frobisher brought Eskimos back from his expeditions to South Baffin Island in 1576–77. Early antiquaries drew comparisons between aspects of American Indian and Eskimo characteristics and those of the ancient Britons. Ethnographical illustration had a great deal of influence on the way that early man was depicted, and Piggott has assembled an unusual series of illustrations of Eskimos, Britons and, most enchanting, Picts and other North British tribesmen.

The fieldwork and recording of monuments by John Aubrey and Edward Lhuyd are set within the context of the day. Our current understanding of monuments such as Stonehenge, Avebury and the Rollright Stones begins with observations made three centuries ago. Thus William Stukeley is seen on the one hand "as a field



Hard facts? — Stonehenge, like Avebury and the Rollright Stones, inspired a "fantastic past for ancient Britain, dominated by fictitious Druids".

strands of research to create a fascinating picture of the intellectual worlds of well-known figures such as Aubrey and Stukeley, and at the same time introduces many lesser known writers from the late sixteenth to the early nineteenth centuries.

The past must be understood within the intellectual constructs of the time, and thus we are not introduced to approaches to prehistory — for the term prehistoric was first used by Daniel Wilson in 1851 — but to a world for which the chronology of early times was outlined in the Old Testament. The early antiquaries are to be seen as individuals undertaking research within

archaeologist of great distinction" and on the other as the man who misled generations "as the inventor and propagator of a fantastic past for ancient Britain, dominated by fictitious Druids" (p. 122). Stukeley's fieldwork at Avebury and Stonehenge is the earliest detailed recording of field monuments; between 1719 and 1724 he spent up to two months at a time surveying them using a theodolite, and making notes and drawings. In a way that we think of as 'modern', Stukeley's concern was to put on record monuments at risk from agricultural and building developments. ►

Piggott sets in context the decline in scientific endeavour in the decades following 1700, and leads us through to the campaigns of barrow-digging in Wiltshire a century later. Some archaeologists today belong to the traditions formulated at that time, and I can feel much sympathy for the "young antiquary, who, from practice, can open a barrow as nicely as you could cut up an apple pie; and from his description, it is done much after the same manner" (p. 156). Archaeologists need to understand how our attitudes to the past

have been moulded.

In this volume Stuart Piggott draws together themes that he has discussed in his studies of the Druids, in his collected essays on antiquarianism, *Ruins in a Landscape*, and in his biography of William Stukeley. Here they are put within an historical and cultural setting that is both accessible and authoritative. □

Graham Ritchie is at the Royal Commission on the Ancient and Historical Monuments of Scotland, 54 Melville Street, Edinburgh EH3 7HF, UK.

Complex stability

Paul Colinvaux

Thinking Green: An Anthology of Essential Ecological Writing. Edited by Michael Allaby. Barrie & Jenkins, London: 1989. Pp.260. £14.95.

In its issue of 23 September 1989, *The Economist* carried a review of the economies of Third World countries. The conclusion was that whether these economies caused disastrous poverty, or wealth like the dragons of Asia, depended only on whether the governments are good or bad, and no mention was made of possible limits to growth on a finite Earth. Writers in the environmentalist 1970s saw things differently. Their concern was for the miserable social consequences of human numbers rising beyond some critical density, or for the fouling of the human nest by industry. Both views are surely right.

But in their anxiety to tell their dreadful truths, the 'green' writers made what Sherlock Holmes called "the capital mistake" of reasoning beyond their data. Their predictions of famines and shortages and limits and catastrophes were not warranted by what we knew. Readers of *The Economist* can therefore ignore them, even as green philosophies begin to mould the politics of the industrial states.

Thinking Green is a first-class anthology of extracts from the sacred texts of the greens, 45 short essays in all. *Silent Spring*, *The Commons*, Malthus, *The Population Bomb* and *Spaceship Earth*; all are here, together with their many supporters and a tiny few of their opponents. The extracts are short, mercifully so in the case of the wearisome establishment rewrites of such documents as *The Brandt Report*, and of *Global 2000* and *Blueprint for Survival*. But the propagandists themselves are generally a very good read. Rachel Carson is superb, standing out even among essays that include Charles Dickens writing on 'Coketown'. This is a book around which to build a seminar for our times.

But the seminar will have to acknowledge much error on the path to truth. Green thinking is riven by longing for a

superorganismic Earth that never was. Green writers declared that the complex was stable and the simple not. Almost no modern ecologist will accept this; mathematics, empirical data and the process of natural selection are all against it. Yet the mystical belief in an organic nature not to be trifled with lingers on. *Global 2000* had in its opening sentence the idiot assertion that the coming world would be "less stable ecologically", and



Thomas Malthus — original thinker.

this was 6 years after Robert May's demonstration of the falseness of complexity-stability theory. The latest reincarnation of a superorganismic Earth is complete with a name from the myths whence it came, 'Gaia', and is allotted two extracts.

The central catch phrase of Malthus, that food supplies grow arithmetically whereas populations grow geometrically, has proved meaningless. Food supplies have often been increased exponentially for prolonged periods. And human beings do not breed like animals, but do so under conscious parental choice subtly affected by cultural values. Malthus himself knew this, and in the extract from his *Essay on Population* declaims against the "viciousness" that follows times when men elect not to marry.

Mercifully, green propagandists now

also realize that they were grievously misled by those who threatened us with a loss of atmosphere as the wages of pollution, so that none of the extracts in the book include ponderings about our running out of oxygen because of the poisoning of plants in the sea. The editor must have chosen his material with care.

With these essential criticisms in the open, the core of *Thinking Green* describes the true dilemmas of our times. The good life requires access to a diverse nature, people in tolerable densities and freedom of choice, qualities that are not so easily provided as is the 'wealth' that economists measure.

The numbers of people can be contained, if we have the patience of generations, as we must. The curse of excessive families is a cultural curse, not a biological one. We do act to regulate our reproduction, even the poorest of us when settled in stable cultures. If we think with the green writers, as I do, that the sooner we bring the numbers down the better, then there is an easy way. Liberate women completely, in culture after culture, so that female lives are fully equal in opportunity to those of men. Can anyone doubt that women so liberated would, on the average, decide to have fewer children?

Some environmental abuses are solvable by law, although neither the laws nor the methods of enforcement have yet been devised. The tragedy of the commons is being played out in the world oceans as Asian ships with 30-mile nets strip out all large living things; world law and warships could correct this, as they did for a different form of piracy long ago.

The misery of peasants in lands once parts of Spanish and British empires comes in part because productive land was taken from them to grow cash crops to support Spanish and English gentry, a way of life that is retained by their local, gentrified successors. In the English heartland of the industrial revolution, subsidized industry — posing as farming — tears down the beauty, albeit artificial, of the countryside to grow sugar beet that no one needs. And the tragedy of sugar production is but a repertory-company version of the grander tragedy being acted out in the Amazon basin.

Politics and law are the only tools that can be used to redress these assaults on decent living. So the greens turn political. Alas they also turn socialist, despite the warnings from the Soviet Union and military dictatorships alike that state control does not lead to the preservation of land or biological diversity. But green thinking is inside politics to stay and the greens may yet be the law-givers. □

Paul Colinvaux is in the Departments of Zoology and Anthropology, The Ohio State University, 1735 Neil Avenue, Columbus, Ohio 43210, USA.

Warming up the Cold War

Daniel S. Greenberg

Techno-Diplomacy: US-Soviet Confrontations in Science and Technology. By Glenn E. Schweitzer. Plenum: 1989. Pp.320. \$22.95.

EARLY in the Cold War, Glenn Schweitzer went to Moscow as the scientific officer at the US embassy. Ever since then he has been deeply involved in one of the most perplexing and disputed aspects of the rivalry between the superpowers — their relations in science and technology, which have been maintained to some degree even during the most uneasy periods of the past 30 years.

To the Reagan administration, visiting Soviet scientists were spies, and research in the Soviet Union was too backward to warrant the interest of US scientists. The Reaganites also rejected as infantile the hope that contact with science in the United States would soften Soviet ideology and promote reforms in harmony with Western interests. Schweitzer, who is now staff director for Soviet and Eastern European Affairs at the National Academy of Sciences, makes the contrary case. The issue of 'warmer' or 'cooler' research relations has at present been rendered moot by the glow of *glasnost* but, as Schweitzer points out, the pendulum has swung many times in the post-war years, and may yet swing again. In brief, he makes three points. First, that scientific exchanges are a poor way of collecting strategically valuable research data, and therefore should not be curtailed because of fears of espionage; second, that although Soviet science and technology lag in many fields, other areas are so advanced that US self-interest dictates collaboration with Soviet researchers; and third, that Soviet scientific and technical contacts have helped "blunt the edge of the US-USSR military confrontation", to which, intriguingly, is added the remark that "in Washington, their importance as a rudder of political stability in the rough turbulence of US-Soviet relations has yet to be adequately recognized".

The book contains an odd mixture of officialese and platitudes ("Capabilities in science and engineering spring from a nation's educational system") and fascinating, but often all-too-brief accounts of Schweitzer's personal experience of Soviet science, technology and higher education. For instance, referring to the academic process of thesis defence, he writes: "I have attended some of these sessions; sometimes they are filled with scientific controversy, while on other occasions they seem almost *pro forma* in

satisfying an academic requirement". And that's it. Elsewhere, he expresses himself forcefully as when he assails one of the ideological hatchet men of the Reagan era, Richard Perle, for his opposition to Soviet-US scientific relations.

There is no indulgence in romanticism about the brotherhood of science to be found here: for example, "Tourism sometimes becomes a principal motivation for American scientists to participate in exchanges. . . . The ballet and the monastery visits can set the timetable". Likewise, a trip to the West is coveted by Soviet scientists because it gives them the opportunity for a shopping spree. Nor does Schweitzer dispute the contention that the Soviets prey upon Western science and technology to redress their own deficiencies in research. He insists, however, that although Western technology has been useful in some instances, it is not the backbone of the Soviet Union's formidable military power. Furthermore, he points out that technologies that originate in the civilian sector — semiconductors, for example — are increasingly vital to military systems, and are increasingly available on world markets.

Throughout, the weight of the argument is that the strategic risks of scientific amity have been exaggerated while the benefits, scientific and political, have been obscured by right-wing misrepresentations. Along the way, Schweitzer challenges the conventional grumping about Soviet visits to US research facilities, pointing out that exchanges are not an easy way of forging ahead in science. He was formerly director of a laboratory of the Environmental Protection Agency, and notes that he "led many attempts to transfer scientific techniques from our institution to others. . . . We had very advanced techniques, the other laboratories needed them to comply with environmental regulations, and we were eager to facilitate the technology transfer". Nonetheless, bureaucratic inertia and differences in equipment and training requirements meant the process was a time-consuming and difficult one.

As I have mentioned, Schweitzer credits Soviet-US scientific relations with an important role in getting the two countries safely through the Cold War. He suggests that the Soviet Union's scientific window on the West contributed to the country's astonishing ideological turn-about by vividly showing up that country's scientific and technological backwardness. And, on behalf of US interests, he pleads for even closer ties amid the turbulence and uncertainties that currently afflict the Soviet Union.

"Let us hope", he writes, "that our two countries are rapidly emerging from the adolescence of scientific nationalism Most importantly, American scientists have known and respected for many

years Soviet scientific colleagues who are now members of Gorbachev's brain trust and who fully appreciate the need to redirect the Soviet scientific enterprise toward peaceful purposes. The likelihood of their doing so depends to a large extent on the US response to the new opportunities for relaxation of military tensions and for scientific cooperation".

This, he concludes, "is the challenge of techno-diplomacy". □

Daniel S. Greenberg is at 3736 Kanawha Street NW, Washington, DC 20015, USA.

Risky business

Lawrence Freedman

Defending Deterrence: Managing the ABM Treaty Regime Into the 21st Century. Edited by Antonia H. Chayes and Paul Doty. Pergamon-Brassey's, McLean, Virginia: 1989. Pp.286. \$32, £20.

THE 1972 Anti-Ballistic Missile (ABM) Treaty remains one of the most solid achievements of arms control. It was a product of the US-Soviet detente of the early 1970s which came to be associated, eventually, to his embarrassment, with



Popperfoto

McNamara giving a Pentagon press conference during the Vietnam war, 1960.

the name of Henry Kissinger. The United States was suffering from the trauma of the Vietnam war and the Soviet Union appeared to be ascendant. Kissinger's hope was to declare a stalemate, create a Soviet stake in the *status quo*, and so prevent a more fundamental challenge to Western interests.

The initiative which eventually led to the treaty can be traced back to the ideas associated with Robert McNamara, US Secretary of Defense for much of the 1960s. While Kissinger understood stability largely in terms of a political arrangement with the Soviet Union, of which an arms-control treaty was an important component, McNamara saw it more as a



THE UNITED NATIONS UNIVERSITY

As a world-wide community of scholars, the United Nations University devotes its work to pressing global problems of human survival, development, and welfare.

• NEW TITLES •

Forests, Climate, and Hydrology: Regional Impacts

Edited by Evan R.C. Reynolds and Frank B. Thompson

This book aims to provide an objective and quantitative assessment of the effects of deforestation or reforestation on climate and hydrology and the regional extent of these effects.

227 pages, 16.5 x 23.5 cm, paperback, US\$20

Agricultural Expansion and Pioneer Settlements in the Humid Tropics

Edited by Walther Manshard and William B. Morgan

The 16 case-studies here examine the driving forces behind the movement of farmers into new lands in the humid tropics.

312 pages, 16.5 x 23.5 cm, paper-bound, US\$20

Research Methods in Nutritional Anthropology

Edited by Gretel H. Pelto, Pertti J. Pelto, and Ellen Messer

Emphasis here is on methods for studying nutritionally related social behaviour and household functioning, determinants and cultural components of food intake, analysis of energy expenditure, and statistical methodologies.

218 pages, 16.4 x 23.9 cm, paperback, US\$30

The Political Economy of Fiscal Policy

Edited by Miguel Urrutia, Shinichi Ichimura, and Setsuko Yukawa

Considers the political obstacles to sound fiscal management through a comparative examination of the decision-making processes of Chile, Colombia, the Republic of Korea, Mexico, the Philippines, and Turkey.

360 pages, 16.5 x 23.8 cm, paperback, US\$30

Developing country prices available. Payment may also be made in Unesco coupons (available from a local Unesco office). A complete list of UNU publications is available upon request.

The United Nations University
APS/NAT
Toho Seimei Building
2-15-1 Shibuya, Shibuya-ku
Tokyo 150, Japan

military arrangement.

During his tenure in office, McNamara had stressed the need for a stable military balance. By this he meant that in a crisis neither side would have any incentive to strike first with nuclear weapons in the hope of denying the opponent a prospect of retaliation. The concept was of mutual assured destruction. Stability was a consequence of a shared vulnerability — when both superpowers might be consumed in a conflagration, neither side would light the touch paper.

Anything that could reduce vulnerability might reduce stability. This included defensive weapons. A 'shield' on its own might seem innocuous — the most moral form of armed force — but it could tilt the balance in a contest between two sides hitherto armed only with swords. During the 1960s the Soviet Union, unconvinced by this notion and anxious to build barriers against all possible attacks against its territory, constructed a first-generation anti-ballistic missile system around Moscow. McNamara did not want to imitate the system because he believed, correctly, that new offensive methods, and in particular multiple warheads, could overwhelm any defensive missiles.

Nonetheless, as he was leaving office he was obliged to authorize the construction of a US ABM system, ostensibly designed to cope with the first Chinese missiles. In 1969 the Nixon administration made this more explicitly anti-Soviet, but faced difficulty in gaining public, scientific and congressional support. The system became dispensable and was sacrificed to achieve the 1972 treaty.

Just over ten years after the treaty came into force it was challenged by the Strategic Defense Initiative (SDI), set in motion by President Reagan in 1983. Reagan and his supporters abhorred the idea of mutual vulnerability. It would be better, the president insisted, to "save lives" rather than "avenge" them. He hoped to find means of destroying offensive missiles during the first stages of their flight paths, from boost phase to the re-entry of the nuclear charges into the atmosphere. From early on it was pointed out that such an effort was in direct contradiction to the ABM Treaty. The treaty prohibited not only the deployment, but also the development and testing of all systems other than fixed, land-based interceptor missiles and their associated radars of the sort that had been in place in 1972 and that could not be modernized, albeit within very strict numerical limits (initially 200 and then later 100 interceptors for each side).

Once the president's will was known the initial reaction from the Pentagon was that, should the research appear promising, the ABM Treaty could be dispensed with if necessary. However, instead of mounting such a direct assault on an arms control treaty, at a time when the

president was claiming to favour arms control in principle, a more indirect approach was adopted, exploring what might be permitted on generous interpretations of the treaty's provisions. The treaty only covered 'strategic' missiles; might it be possible to develop a system to deal with 'tactical' missiles in Europe? There was overlap between anti-satellite and anti-missile weapons; might the former be used as a cover for the development of the latter? The treaty limited the testing of 'components', but what might be done with sub-components? There was a complicated provision concerning the testing of systems based on 'new physical principles'; despite uncomplicated and apparently unambiguous language to the contrary, could this be construed as allowing the development and testing of anything involving new technologies?

This assault on the treaty was complicated by the Soviet response to the SDI. Instead of simply saying that if it went beyond research it would be prohibited, Soviet spokesmen insisted that even research was not allowed. Later they proposed an extended mutual commitment to staying within the treaty's terms, but by this time the main issue had become the Reagan administration's broad interpretation of the terms. The Soviet position was not helped by doubts over its own compliance. Matters became more complicated with proposals to get out of the impasse by considering agreed experiments that could be undertaken without undermining the treaty completely.

Those who prefer the known strengths of the treaty to the doubtful promise of SDI have thus been forced to pit technicality against legality. The number of people truly competent to do this are few and most of them have contributed to the book under review. All of the main areas of controversy are addressed. The conclusion, not surprising from the book's title, is that the purpose and constraints of the treaty are far clearer than those who have sought to undermine them would have us think. I do not believe that the opposing view could have assembled anything approaching the same weight of argument and expertise, even though it notionally remains that of the Bush administration.

Defending Deterrence is not an easy read, and it would have been refreshing to have had a more critical look at the concept of deterrence which is being defended. But the book is comprehensive, authoritative and lucid. Should negotiations on the future of the treaty gather pace during the coming months it will be hard to find a more useful work of reference; if the two governments decide that it is all too difficult, then Chayes, Doty and their contributors will help explain why. □

Lawrence Freedman is in the Department of War Studies, King's College London, Strand, London WC2R 2LS, UK.

For medicinal purposes

Alan Glynn

Plagues and Politics: The Story of the United States Public Health Service. By Fitzhugh Mullan. Basic Books, New York: 1989. Pp. 223. \$26.95.

THE United States Public Health Service (PHS) is the direct and legitimate descendant of the United States Marine Hospital Service (MHS), which was born, by Act of Congress, in 1798. What then is this centennial volume, numerate readers may ask? The centenary involved is that of the 'Commissioned Corps', established (or at least formalized) by another Act in 1889, which formed an élite of medical men until the pressure of the Second World War forced the entry of non-medical scientists, engineers and women.

Mullan deals with the whole of the PHS's activities, but it is the Commissioned Corps who are the good guys. 'Plagues' are the leading baddies, though 'Politics' are not always far behind. In 1870 a reforming Act recommended central control of the MHS by a supervising surgeon in Washington, but stipulated that he must be drawn from civil life. This blocked the appointment of the man who had contributed most to the report on which the Act was based, John Shaw Billings, arguably the best Surgeon General the PHS never had. The reason for this restriction is not made clear. Anti-militarism is less likely than simple politi-

cal favouritism. Billings's successful rival, John Woodworth, had been the senior medical officer under Sherman.

In an age when there was blatant outside interference in civil-service appointments, Woodworth copied the army model for appointments and later introduced stiff professional examinations for entry to the MHS. The title Supervising Surgeon changed imperceptibly to Supervising Surgeon General, and finally to Surgeon General. Meanwhile, Billings was working in the office of the Surgeon General (of the army), and was building up what became the greatest medical library in the world and introducing a new idea, the *Index Medicus*. Ironically, both were transferred to the PHS in 1956.

These early rivalries set the tone for the next century. Successive Surgeons General had to contend with civil-service jealousies because pay in the PHS was better. By restricting entry to the Commissioned Corps to physicians, they made it difficult to recruit scientists of quality, so handicapping the work and limiting its scope. Paradoxically, as Mullan points out, they also fought numerous proposals to develop aspects of public health not because they were undesirable but because they might have produced competing agencies.

Plagues and Politics is not just a picture book (although the pictures are very good). It gives a very readable account of the battles the PHS has fought, on the whole successfully, against disease, the US treasury and public health reformers in and outside government. Against disease the PHS's record is extraordinarily good. Even the notorious Tuskegee

project, if judged by the ethics of its time, was at least well meant. Goldberger's demonstration of the nutritional basis of pellagra and Stiles's work on hookworm are just two examples of many successful investigations and campaigns.

In a different category comes Lumsden's pioneering work on rural sanitation. Unfortunately, his views on federal help for states differed from those of Surgeon General Cummings, and he was eventually 'exiled' to New Orleans. Indeed, one of the alleged merits of the Commissioned Corps was that its members could be posted anywhere. This also meant overseas, and the PHS's world-wide experience of disease problems must have influenced Surgeon General Parran, who played a large part in the creation of the World Health Organization. Although Mullan makes several references to the Centers for Disease Control, the epidemiological nerve centre which grew out of the Program for Malaria Control in War Areas, he does not adequately convey its full importance.

Early political battles were mostly about money or federal-state relations. The later struggles for organizational power are more difficult for an outsider to understand, because they involve many interests, to say nothing of the acronyms for numerous and sometimes ephemeral agencies. Not infrequently, politics and disease were both involved. When Kinoun was sent to deal with the first outbreak of plague in the United States he was vilified by the Governor of California and other local interests, whose reactions were similar to those of some local authorities in England when cholera arrived in 1832. After the Second World War the changed views of government and public on public health inevitably brought changes to the traditional PHS. After some anxious moments in the 1970s, when only an acting Surgeon General was appointed and all authority was vested in the Assistant Secretary for Health, the PHS is now reasonably settled as part of the Department of Health Education and Welfare. In his campaign against AIDS, the former Surgeon General, C. Everett Koop, has upheld the PHS's tradition of attacking disease problems where it thinks necessary, whether or not governments are keen to follow.

Perhaps *Plagues and Politics* should be looked on as a forerunner to some future monumental work on the PHS that will answer all the unanswered questions. But such a book is unlikely to be as readable. Perhaps, too, someone will write a similar history of public health and government in Britain, comparing the Surgeon General with the Chief Medical Officer. It would be fascinating and instructive. □

Alan Glynn, Central Public Health Laboratory, Colindale, London NW9 5HT, UK, retired last year as director of the laboratory.



Socially transmitted enthusiasm — urban support for the campaign to control syphilis in Chicago during the late 1930s (reproduced from the book in review).

Rules of conduct

John A. Campbell

Benefits and Risks of Knowledge-Based Systems. Report of a Working Party of the Council for Science and Society, chairwoman Margaret Boden. *Oxford University Press: 1989. Pp. 76. Pbk £6.95, \$14.95.*

ARTIFICIAL intelligence (AI) is in the news less often now than it was in the early and middle 1980s. At present there is nothing quite like the world-wide interest in the Japanese 'Fifth Generation' plan or in the novelty and first commercial successes of expert systems that produced all the previous attention. But AI research operates rather like a volcano: there is constant activity underground, which becomes visible only during eruptions. It is likely that neural nets will cause the next eruption, and in the not-too-distant future.

Among AI researchers there is the saying "AI exports its successes". What this usually means is that the wider public that makes use of the successes no longer sees them as part of AI (the languages LISP and Prolog are examples). Now expert systems seem to be going in the same direction, like lava that is cooling down

into easily-handled pieces of stone.

In some quarters, this comfortable trend towards cooling or domestication is even seen as not being fast enough. In the ESPRIT programme, which is an EEC response to the Fifth Generation plan, calls for proposals for research and development projects regularly contain subject-headings such as "metrication and validation of knowledge-based systems". The implication is that computer programs built around knowledge (essentially the rule-based knowledge that is the distinguishing feature of a typical expert system) should eventually be metricated, validated and domesticated with no more trouble than for book-keeping or ticket-issuing programs.

To do the writers of the ESPRIT material justice, they appreciate that they are expressing a hope rather than a certainty. It is unfortunate that this appreciation is not more widespread, especially among people who use or rely on expert-systems technology. The book reviewed here illustrates some of the issues by highlighting the present limitations of the technology, and makes suggestions about how to deal with them. It is welcome because the suggestions are modest and reasonable — for example, that the systems should be treated not as oracles but as assistants to people performing tasks that require specialized knowledge, and that it should be possible for a user to examine the knowledge represented in them. But there is a risk that the book's modesty and low profile will mean that its arguments go unnoticed. Here is a rare case of a book's packaging being too sober.

The special character of knowledge-based systems (KBS) follows from the fact that knowledge is multi-faceted, open-ended, ambiguous and even scruffy, while the functions and specification of (say) book-keeping programs are not. Admittedly there are areas in which the relevant knowledge is well-behaved and can be given the book-keeping treatment, as in some of the most commercially successful expert systems, but these examples do not mean that all conceivable or desired KBS must have the same properties. One possible criticism of the book is that, by using the terms 'expert system' and 'knowledge-based system' almost interchangeably, it misses a chance to sharpen this point.

Clearly a KBS is founded on knowledge, in whatever representations are available to hold it. An expert system obviously has something to do with expertise, but its original meaning (most of which still survives, particularly in the United States), was tied quite tightly to one single representation for knowledge: *rules*. The most interesting KBS may use rules, but other representations such as frames (collections of information that belong together in some particular con-

text) and inheritance nets (hierarchical structures in which specific items inherit properties from more general items, of which they are particular instances) are used at the same time. This reflects the actual organization of expert knowledge much better, and extends the range of AI applications beyond what was available in the early 1980s. But it also emphasizes the difficulty of predicting all the outcomes and pitfalls when some particular body of knowledge that goes beyond mere rules is represented and used inside a KBS. In this respect, the book gives a fair summary of the issues that deserve to be confronted, even though simple rules are the only examples of knowledge that it uses. To get the most out of the case that it makes, the reader should repeat regularly: "There is more to knowledge than rules".

The fundamental reason for the mixture of cautions and actions that the authors' collective for this book recommends is that a KBS is unlikely to use the same chain of reasoning that a human would use to solve the same problem. The methods of inference in a KBS are local approximations to our own reasoning for parts of problems — sometimes optimistic approximations (for example, the psychological evidence that we do any reasoning with the help of logic as a mathematical logician would describe it is very thin). But there is no general AI method that imitates our perception and treatment of a reasoning problem as a whole. This gives force to the recommendation that a KBS should be used as a human's assistant and not (unless provably domesticated) trusted as an autonomous agent, especially where human life and safety are concerned. An assistant who suggests initiatives should always be capable of explaining them in terms that a human senior partner can understand and accept. This concentrates attention on the ability of a typical KBS to explain its actions, and here there is still plenty of room for improvement. Explanation, or lack of it, is also likely to be the main weakness of neural-net systems, despite their strengths in other respects. Therefore, the book's relevance will not be changed by the arrival of neural nets on the commercial scene, if this indeed causes the next eruption of publicity for AI.

Relevance and reasonableness alone are not enough to make a political impression, alas. Subversion of the established (commercial) order is a better way in 1989. If the author's subversive recommendation that a "human or group of humans ... should take moral and/or legal responsibility for the human consequences of any malfunction" of any KBS does not stir up some debate, then nothing will. □

John A. Campbell is in the Department of Computer Science, University College London, Gower Street, London WC1E 6BT, UK.

THE MOTOR THEORY OF LANGUAGE ORIGIN

Robin Allott

How in evolutionary terms did humans acquire language? Language cannot have sprung suddenly into existence; it must have had a precursor. The motor theory is that the precursor was the motor control system, the complex organisation of the brain for movement. The book brings together evidence from two major areas of research, linguistics and the neurology of human perception and action.

£8.50 THE BOOK GUILD
25 High St., Lewes, Sussex